CLASSIFICATION MANAGEMENT AND USE IN A NETWORKED ENVIRONMENT: THE CASE OF THE UNIVERSAL DECIMAL CLASSIFICATION

By

Aida Slavic-Overfield

Thesis submitted for the Degree of Doctor of Philosophy at the University of London

University College London

London, April 2005

ACKNOWLEDGEMENTS: This research project was realized through the participation in the UDC Revision Committee work during the period 2000-2005. This would not have been possible without the help and support of the Editor in Chief of the UDC, Prof. I. C. McIlwaine, who I would like to thank for supervising and supporting my work.

I would like to express my gratitude to my colleague and friend Maria Ines Cordeiro for sharing the excitement and enthusiasm of improving standards for classification management and exchange and for inspiring and challenging discussions valuable to this research. I am also grateful to Dr. Gerhard Riesthuis of the University of Amsterdam for exchange of information and his thoughts on UDC MRF. Thanks are also due to Prof. Jadranka Lasić-Lazić of the University of Zagreb for encouragement and support that made this doctoral study possible.

Finally, I would like to express my heartfelt thanks to my husband, Chris Overfield, who graciously immersed himself in the world of library classification and escaped unscathed. He kindly donated his programming skills in several research projects involving classification that were beneficial and inspired this doctoral research.

ABSTRACT

In the Internet information space, advanced information retrieval (IR) methods and automatic text processing are used in conjunction with traditional knowledge organization systems (KOS). New information technology provides a platform for better KOS publishing, exploitation and sharing both for human and machine use. Networked KOS services are now being planned and developed as powerful tools for resource discovery. They will enable automatic contextualisation, interpretation and query matching to different indexing languages. The Semantic Web promises to be an environment in which the quality of semantic relationships in bibliographic classification systems can be fully exploited. Their use in the networked environment is, however, limited by the fact that they are not prepared or made available for advanced machine processing. The UDC was chosen for this research because of its widespread use and its long-term presence in online information retrieval systems. It was also the first system to be used for the automatic classification of Internet resources, and the first to be made available as a classification tool on the Web.

The objective of this research is to establish the advantages of using UDC for information retrieval in a networked environment, to highlight the problems of automation and classification exchange, and to offer possible solutions. The first research question was is there enough evidence of the use of classification on the Internet to justify further development with this particular environment in mind? The second question is what are the automation requirements for the full exploitation of UDC and its exchange? The third question is which areas are in need of improvement and what specific recommendations can be made for implementing the UDC in a networked environment?

A summary of changes required in the management and development of the UDC to facilitate its full adaptation for future use is drawn from this analysis.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AAT	Art and Architecture Thesaurus [KOS]
AMICO	Art Museum Image Consortium (data dictionary)
ASCII	American Standard Code for Information Interchange
BBK	Bibliotechno-Bibliografska Klasifikacija (Soviet Library and Bibliographic Classification) [KOS]
BSI	British Standards Institution
BUBL	Bulletin Board for Libraries
CORC	Cooperative Online Resource Catalog
CC	Colon Classification [KOS]
CDU	Italian, Spanish, French, Portuguese equivalent for abbreviation 'UDC'
DAML	DARPA Agent Markup Language
DCMES	Dublin Core Metadata Element Set
DCMI	Dublin Core Metadata Initiative
DDC	Dewey Decimal Classification [KOS]
DESIRE	Development of a European Service for Information on Research and Education
DK	German equivalent for the abbreviation 'UDC'
E&C	Extensions & corrections to the UDC
EAD	Encoded Archival Description
EASEL	Educator Access to Services in the Electronic Landscape
Ei	Engineering Information thesaurus [KOS]
eLIB	The Electronic Libraries Programme
ETH	Eidgenössische Technische Hochschule
ETO	Hungarian equivalent for the abbreviation 'UDC'
FATKS	Facet Analytical Theory Knowledge Structure
FID	International Federation for Documentation & Information
FRBR	Functional requirements for bibliographic records
GERHARD	GERman Harvest Automated Retrieval and Directory
GUI	Graphical User Interface
HTML	Hypertext Markup Language
IFLA	International Federation of Library Associations
IR	Information retrieval
KO	Knowledge Organization
KOS	Knowledge Organization System
LCC	Library of Congress Classification [KOS]
LCSH	Library of Congress Subject Headings [KOS]
LIS	Library and Information Science
LOM	Learning Object Metadata, known also as IEEE LOM
MARC	Machine Readable Cataloging
M2M	Machine to machine

NEBIS	Netzwerk von Bibliotheken und Informationsstellen in der Schweiz
NIDR	Networked Information Discovery and Retrieval
NISS	National Information Services and System
NKOS	Networked Knowledge Organization Systems/Services
NLM	National Library of Medicine classification [KOS]
OCLC	The Online Computer Library Centre
OIL	Ontology Interface Layer
OMNI	Organising Medical Networked Information
OPAC	Online Public Access Catalogue
OWL	Ontology Web Language
PSI	Published Subject Identifier
RDF	Resource Description Framework
ROADS	Resource Organisation and Discovery in Subject-based Services
SDI	Selective Dissemination of Information
SGML	Standard Generalized Markup Language
SKOS	Simple Knowledge Organisation System
SOSIG	Social Science Information Gateway
TEI	Text Encoding Initiative
UDC	Universal Decimal Classification [KOS]
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
XML	Extensible Markup Language
XTM	XML Topic Maps (XML specification for ISO/IEC 13250:2000 Topic Maps)
VDEX	Vocabulary definition exchange
W3C	World Wide Web Consortium
WWLib	The Wolverhampton Web Library
WWW	World Wide Web
WAIS	Wide Area Information Server
WAIS/WWW	Nordic WAIS/World Wide Web

CHAPTER ONE: RESEARCH BACKGROUND AND REVIEW OF THE LITERATURE

1.1 Research fields and issues

The 1970s were witness to a decline in the use of classification as a tool for information indexing and retrieval, in favour of alphabetical subject indexing languages and free-text searching (Silva, Santos & Lopes, 1989). This trend continued through the 1980s and 1990s, resulting in a number of libraries and documentation centres migrating to subject-alphabetical indexing languages, or abandoning indexing languages altogether.

The decrease in the use of classification during the period, brought about by library automation practices, followed different patterns. In the majority of Anglo-American libraries, where automation first developed, classification was mainly used for shelf-arrangement and not IR, and facilities for searching and browsing classifications were not standard in vendor library systems. Although the situation was intensely criticized throughout the 1980s, it still remains a problem at the time of this research (Gnoli, Ridi & Visintin, 2004; Long, 2000; Borgman, 1996; Cousins, 1992; Hildreth, 1991, Howard, 1990; Wajenberg, 1983; Svenonius, 1983; Cochrane & Markey, 1985; Markey Drabenstott, 1986, 1987).

In a multilingual/multiscript European library environment, however, classification had greater value in IR, which led to two distinct practices. The first was typical for countries where UDC was predominantly used for IR and not for shelf arrangement such as Austria, France, Germany, Portugal and Scandinavian countries. Here, the classification suffered a decline in use in favour of keyword searching, subject-headings, thesauri or other solutions offered by library system vendors. The second practice was relevant for Eastern European countries and a few libraries in Western Europe, where classification was fully used in bibliographic services for shelf arrangement, IR and bibliographies, and was paramount for information exchange. Library systems in this environment were often developed in-house, in local library networks or at a national level, such as is the case with the Czech Republic, Slovakia, Hungary, Romania, Slovenia and Croatia. In these countries, there was no noticeable decline in the use of classification. This practice was, however, largely outside the mainstream of library automation and of classification research and remained isolated.

User studies and research on problems in subject searching of library online public access catalogues (OPACs) were mostly focused on the Dewey Decimal Classification (DDC) and Library of Congress Classification (LCC) systems. Research on OPACs using UDC, typically for European library practice, was not so prevalent in the 1990s and there was a general lack of awareness that a more standardized approach was necessary for classification management. Success stories of the use of classification in OPACs, such as those provided by Buxton (1990), did not manage to attract attention to the general and widespread problem of the poor practice of classification in library systems. Although library classification shares the same environment, the research on Colon Classification (CC), DDC and LCC automation was not sufficiently aligned to the UDC in relevant points of standardization.

The fact that classification systems, by and large, were not fully employed by library systems inhibited their adaptation for online use and excluded them from mainstream IR development. As a consequence, standards for the management of classification data in automated systems have been very slow in development.

The growth of the Internet, in the period 1990-2004 and the new information and communication environment brought a positive change in attitude towards classification. The infrastructure necessary for the exchange of classification vocabularies, however, has still not been put in place. As a result, classifications such as UDC have only been partially utilized in information resource discovery. In addition, a lack of input in the classification machine-readable source data from publishers of classification systems, and a lack of expertise among those who are supposed to use it continue to be the main reasons hindering the use of classification in IR (Slavic & Cordeiro, 2004, 2004a).

The networked environment endorses general and system independent solutions and imposes the philosophy of an open information space in which the same technological vehicle is used to transport many different kinds of content. Alignment with this philosophy is where the majority of work needs to be done so as to enhance the standardization of management, use and exchange of classification in a networked environment.

This chapter establishes a framework for analysing classification in a networked environment. It begins by explaining the historical context of the UDC automation and continues by examining the general issues of using classification systems in IR. Finally, it looks into the rôle classification may play in knowledge organization and resource discovery on the Internet.

Starting with most specific topic - the UDC - and moving towards general issues of the networked environment, this section will give an overview of variables that directly influence the use, management and consequently the exchange of classification systems.

1.2 UDC automation

UDC was created in 1895 to index and organize 'the greatest bibliography ever put together that would include everything published, anywhere in the world, in any language, at any time'.[1] Although based on DDC, which is a typical library classification, the UDC was unique amongst the library classifications of its time in that it was not designed for library shelf arrangement. To serve its purpose in this universal bibliography, the UDC was designed as an indexing language proper. Its vocabulary was organized relationally into different sets of hierarchies with syntax rules to enable unlimited re-use in different combined subject expressions suitable for detailed indexing, easy expansion and wide application.

In spite of the fact that UDC was created too early for any mechanisation to be anticipated, either in the maintenance of the classification or in its application, there is a significant amount of modern thinking and IR expertise behind it. For instance, the UDC notation is designed to carry information about both hierarchy and synthesis, which was a novelty at a time when the notation of other library classifications was created with a single concern in mind: book labelling for shelf arrangement. The UDC data representation is formalised to a level unusual for library classification and is made less dependent on the form in which schedules are printed or displayed. Consistency in representation and use of symbols make the

[1] The Universal Bibliographic Repertory (Répertoire Bibliographique Universel), as it was called by its creators Paul Otlet and Henry LaFontaine, was a documentation project on a global scale. This bibliographic undertaking ultimately failed after producing a repository of no less than 11 million records and leaving as its main legacy the UDC system and its owner up to 1992 - *International Federation for Information and Documentation* (FID) (Dubuc, 1973; Rayward, 1994). FID ceased to exist at the end of 2000 (Horton, 2003).

system rules easy to mechanise.

The most specific feature of the UDC is its extensive analytico-synthetic function, which means that elements of the classification can be freely combined on several levels. The synthesis is based on two main structural qualities *(a)* defined categories of hierarchically organized vocabulary (sometimes structured according to fundamental facet categories, sometimes a set of pragmatically chosen classes of vocabulary) and *(b)* flexible syntactic rules using symbols that allow the combination of concepts from different categories or from different disciplines.

Being suitable for both a hierarchical organization of subject and detailed indexing, UDC has been used for the following purposes:

i) systematic arrangement of objects/resources in:
   - library shelf arrangement
   - organization/presentation of information resources online
   - organization of objects (realia)
ii) systematic arrangement of surrogates for IR and discovery in:
   - bibliographical databases, bibliographies, library catalogues
   - digital collections, full text databases (with embedded metadata)
iii) alerting services/selective dissemination of information/current awareness bulletins
iv) creation or management (mapping) of other controlled vocabularies

The first two application areas can be seen to be constant and stable. The use of the UDC in creating mappings between different indexing languages has gradually increased in importance since the 1990s, whereas its use in alerting services has significantly decreased. All of the above-mentioned applications are usually automated nowadays and their success depends on the management of classification as a single tool that underpins different functions. The issues of functions of the UDC that depend on the automation of its structure are the very focus of this research and will be addressed in Chapters Four, Five and Six.

The UDC was used in an online environment as early as the 1960s. The field of UDC automation can be viewed as three distinct and functionally interdependent parts: automation of the UDC schedules, automation of data retrieval from the UDC and automation of the classification process. The **automation of UDC schedules** focuses on two main areas: enabling the maintenance, development and distribution of the classification and enabling the construction of classification tools. This is the most recent development in automation. The **automation of the UDC retrieval process** is tied to information systems and is the oldest and most studied area. The third part deals with the automation of the classification process itself, i.e. **automatic indexing *using classification*,** which in theory, is logically dependent on both previously mentioned areas. In practice, there is little interaction between these fields of automation and in the history of UDC automation these areas seem to follow separate and independent paths.

1.2.1 Automation of the UDC schedules

A policy of continuous improvement and subject extension of the UDC was especially encouraged after World War II when the UDC was exposed to large-scale revision efforts (Dahlberg, 1971, 1971a). As the result of management policies by FID, towards the end of the 1960s UDC had an elaborate revision system in place: a system of proposals (P-notes), a system of authorised changes (Extensions & Corrections) and a controlling body of 25 members in FID Central Classification Committee (FID/CCC) coordinating work of National and International Subject Panels (Foskett, 1973: 26). This practice proved too slow and inefficient in the revision of existing schedules and also in keeping pace with science and technology, but even more importantly, it was too slow in distributing changes to end-users.

According to Foskett (1973) the pressure for revision came from two external feasibility studies. The first was commissioned by UNESCO in 1961 and reported by Kyle (1961) and Vickery (1961) who suggested that major restructuring of the UDC towards a faceted classification would be needed.

In 1967 the UNISIST and the *International Council of Scientific Unions* commissioned a comparative study of indexing languages for the development of the information network from the *Aslib Research Department* team, presided over by B. C. Vickery. Although UDC was the 'least unsatisfactory' and was judged to be 'less faulty' than other systems, Vickery proposed that a new classification was needed for the purpose of information exchange (Foskett, 1973). This criticism triggered action from FID in the seventies and efforts were made towards finding solutions for both UDC restructuring and changes in UDC management and maintenance.[2]

Lloyd (1971), the then head of the classification department in FID, claimed that the UDC was in need of 'drastic treatment and more efficient organization' that would work faster than the 'hyper-democratic' revision procedure in place. Wellisch (1971) also argued that the existing procedure was hindering the real progress of revision and that complete restructuring of the organization was needed. His proposal was that decision-making should be transferred from FID/CCC to an editorial board consisting of four or five members working under a chief editor. Revision would be recorded on magnetic tape, which would be copied and distributed to national bodies. His suggestion that a 'medium' edition would satisfy the need of most users was, for instance, also supported by Lloyd (1972).

The idea of the automation of the UDC schedules had been discussed as an appropriate solution to the revision and management of the system since the 1960s (Rigby, 1971). The slow process in automating the classification coincides with an overall crisis of traditional bibliographic classifications that was inflicted by library automation in the 1970s and 1980s.

Numerous projects and studies of UDC mechanization and its use in IR reported in the 1960s and 1970s revealed weaknesses in the classification's content and structure (First seminar on UDC Mechanisation, 1969; Second Seminar on UDC Mechanisation, 1971; Proceedings of the International Symposium UDC in Relation to other Indexing Languages, 1976).[3]

In the 1970s, UDC was suggested by FID and seriously considered for the conversion of information retrieval languages i.e. as a 'switching language' for the, then new, UNISIST programme. At this point the UDC structure and vocabulary was

---

[2] Vickery's analysis was not considered to be entirely fair towards the UDC and was often criticised in the literature in subsequent decades (Newcombe, 1972; Santoro, 1996). Gilchrist, for instance, expressed doubt about the motives of the UK Classification Research Group (CRG) towards the UDC at a time when they were attempting to attract funding for their own projects (Gilchrist, 1992).

[3] FID (financed by UNESCO) organized three seminars devoted to the use of the UDC in an online environment: *First Seminar on UDC in a Mechanized Retrieval System* conducted by R. R. Freeman and Pauline Atherton, Copenhagen, 2-6 September, 1968, *Second Seminar on UDC in Mechanized Information Systems* conducted by R. R. Freeman, Frankfurt, 1-5 June, 1970 and *International Symposium "UDC in Relation to other Indexing Languages"*, Herceg Novi, Yugoslavia, 28 June-1 July, 1971.

closely studied so as to find out whether or not it would be suitable for such a function (Gilchrist, 1972; Lloyd, 1971; Santoro, 1996; UNISIST study report on the feasibility of a world science information system, 1971). Although the idea was abandoned in favour of creating a new classification, this influenced many activities and proposals for UDC restructuring.

Many of the objections to the classification from that period were a consequence of the extensive and decentralized expansion of the classification, which led to the overlapping of concepts that had been added to the existing 19th century semi-enumerative structure. At the beginning of the 1980s the full edition of the classification had exceeded 200,000 classes, as a result of the work of many national and international subject panels. The work was not well coordinated and the plethora of concepts, and the incoherent manner in which they were added to the system, contributed to the difficulties in management and mechanization.

By the end of 1980s there was already a collection of proposals. Atherton and Freeman suggested better recording of explicit syntax rules and strict formalisation of notation (Freeman, 1969; Freeman & Atherton, 1969). Perreault (1969) and Caless (1969) recommended better guidelines for content analysis and the creation of complex numbers. Lloyd (1972) proposed vocabulary and syntax improvement. Wellisch (1976) suggested that more precise revision rules would be needed with systematic and consistent principles for the recognition and application of characteristics of division within main classes. He considered the then existing guidelines "UDC revision and publication procedure" published in 1968 to be unsatisfactory.

Rigby (1971) suggested adding better cross-referencing facilities and subject-alphabetical indexes. Dahlberg (1971, 1971a) emphasized that UDC numbers needed to be such that they could be "fed into a computer in wholly unchanged form". She also suggested the following measures: the abolition of the meaningless decimal point as this would reduce the size of numbers and changing facet indicators for main auxiliaries into letters following the Atherton and Freeman solution. Kyle's (1961) and Vickery's (1961) suggestion for system faceting and better exploitation of UDC’s synthetic features and facet analysis was later revisited by Dahlberg (1971a) and Neelameghan (1976).

Dahlberg's (1971, 1971a, 1998) proposals for changes to the schedules themselves were so fundamental that she eventually decided to concentrate on a new classification scheme instead of improving the UDC. Her ideas about faceting the UDC were, however, well thought through and were linked to the practical issue of the automation of the schedules. She, for instance, correctly realised and suggested that if UDC were to be fully faceted, then it would need to have schedules structured into three parts for the system to be automated: *(a)* UDC tables - containing simple concepts; *(b)* UDC codes - containing compound numbers; *(c)* UDC Index - possibly in the form of a thesaurus (Dahlberg, 1971, 1971a, 1998).[4]

At the beginning of 1983, the FID commissioned an external management study to find a better way of managing the UDC. The study results were published in 1984 and reported the following positive outcomes: the proposal of a new UDC management structure that later helped in handling the problem of computerising the UDC maintenance process, undertaking a strategic review of revision needs, and taking the necessary steps to change faults in the classification structure (Gilchrist, 1992, McIlwaine, 1997a). The UDC Management Board was created in 1986 to carry out the changes and in 1989 it decided to set up a *Task Force for UDC System Development*, with international membership,[5] which in a short period of time produced a strategy paper that offered a more concrete plan for future long-term development of the UDC.

The changes of the revision and management of the UDC proposed by the Task Force were suggested at a time (i.e. 1989/1990) when it was logical to assume that the classification would be stored in a database. In order for the whole project to be feasible, it was proposed that the UDC schedules be reduced to a size most commonly associated with medium editions (as was suggested by Wellisch and Lloyd two decades earlier). Of all the proposals put forward by the Task Force, the

---

[4] The commitment to develop UDC as a faceted scheme was made official by an FID/CCC Extraordinary meeting in 1976. Document C75-35 was released which stated that UDC is to be transformed into a fully faceted scheme based on a methodology put forward by A. F. Schmidt and J.-H. de Wijn (Ścibor & Shcherbina-Samojlova, 1990). Due to the size, ambition and cost of the project, this plan never became operational but it has influenced future incremental and gradual transformation of the schedules towards becoming a faceted classification (McIlwaine, 1993, 1997a; McIlwaine & Williamson, 1994). Very few elaborate suggestions for the improvement of the UDC, and certainly none that are mentioned here, have been implemented for the entire UDC classification by 2005.

[5] The members of the UDC Task force were: A.-R. Haarala, Finland; H. Jobst , Austria; I. C. McIlwaine (chairman), United Kingdom; Gerhard Riesthuis, The Netherlands; Nancy Williamson, Canada. The observers were A. Gilchrist (United Kingdom), David Strachan, FID (Task force for UDC system development: final report, February 1990).

one that most affected UDC use online was the abolishment of the so-called '10 year rule', which was deemed necessary in order to facilitate the revision task (Task force for UDC system development: final report, 1990; McIlwaine, 1990).[6] This allowed vacated UDC numbers to be immediately re-used to denote different concepts and, hence, put more requirements on the appropriate identification of UDC numbers in the process of classification exchange and update.

The conclusions of the Task Force suggested that work needed on maintenance required more financially viable publishers than the FID in its, then, internationally weakened form. On the 1st January 1992, FID transferred its UDC rights to the newly formed Consortium of publishers that was established as a non-profit institution, legally based in The Netherlands and registered as a Dutch foundation (Strachan & Oomes, 1993, 1995). Following the change in ownership, the UDC revision process and decision-making was centralized and transferred to the revision committee consisting of two to three members and an editor in chief. This made the process of revision more rapid and reduced maintenance costs.

In 1993, Gerhard Riesthuis and David Strachan created a database to hold UDC schedules. A text file of the *British Standards Institution* (BSI) medium edition (BS 1000M) of 1985 was used as the basis and imported into the database together with subsequent extensions and corrections.

The automation of schedules has facilitated the maintenance, publishing and use of the classification. Since 1993, the master file of the UDC has been updated and issued annually in a file form and the revision process has been accelerated. The extent and the way in which it has, in reality, affected the revision, maintenance and use in the period 1993-2005 will be further analysed in Chapters Seven and Eight.

1.2.2 Automation of UDC retrieval

UDC was the first library classification designed for detailed indexing and the first to use a notational device with formalism and a grammar paramount for the automation of the retrieval process (Bhattacharyya & Ranganathan, 1978). Its notation is independent of the textual schedule display and is formalised and

---

[6] This rule prevented once cancelled UDC numbers to be re-used in the period of 10 years, which would give enough time to libraries to reclassify and remove the UDC numbers prior to their being re-introduced with a new meaning.

expressive with respect to semantics and syntax.

Capable of supporting greater specificity in indexing, the UDC was adopted by technical and scientific special libraries in research institutes worldwide.[7] These libraries were the first to engage in the automation of information retrieval, thus UDC became the first classification used in automated document searching. Rigby's comprehensive and detailed annotated bibliography on the issue: *Automation and the UDC, 1948-1980* (Rigby, 1981) provides a reliable source of references for the given period. From 1980 to 1993 the use of UDC in IR can be followed through research papers and reports. This period was characterised by a general decline in the use of classification in IR.

1.2.2.1 UDC in IR 1948-1980

Up to 1960 several UDC mechanisation attempts were made using edge notched, peek-a-boo and punched cards. The introduction of early computers and tabulators in the 1960s started the new phase (Batty, 1981:141). These first early mechanisation attempts were marked by projects in the USA, France, Belgium, DDR/FDR, and Hungary (Rigby, 1971). Early computers, however, had very limited functionality when it came to character management and display, and early difficulties with UDC mechanisation were related to these shortcomings (Dubuc, 1973). The following were often reported problems: differences in the length of UDC numbers; variations in the number of UDC classmarks assigned to a single document; the number of symbols used and the non-existence of a fixed order for the appearance of the symbols. This early phase of automation has its historical value and its rôle in popularisation of the UDC but did not greatly influence the future of the field.

Real automation started towards the end of 1960s and was characterised by numerous projects and intensive research activities. The most significant developments for this period were practical IR and Selective Dissemination of Information (SDI) applications and truly operational systems. Rigby listed around 80 institutions worldwide that created computer applications, catalogues and services based on the UDC and published reports on their work (Rigby, 1981: 138).

---

[7] UDC use worldwide is described in detail in Chapter Seven.

The most significant legacy from this period came from two projects: the first was led by Freeman & Atherton and the second by Caless. Their work was based on the understanding of both the UDC system and IR, and was carried out with technology at the leading edge of its field at the time. These achievements were comparable with any present IR system using classification. The results were far reaching and were often referred to whenever there was a need to defend the rôle of classification in IR (Cochrane & Markey, 1985; Cochrane & Johnson, 1996; Koh, 1995; Pollitt, 1997).

UDC automation by Freeman and Atherton - Robert R. Freeman and Pauline Atherton (later Cochrane) were sponsored by the American Institute of Physics (AIP), to design and demonstrate an IR system using UDC. The objective of the AIP/UDC project (1961-1968) was to create mechanized search and display functions and evaluate the ability of UDC notations to present relevant search results in a useful display format. The system consisted of *Master* and *Inverted* files using UDC notations as descriptors and allowed the following functions: searching of truncated indexes, the possibility of searching for the beginning of a string with a defined ending irrespective of the middle part of the index, and the possibility to search using Boolean statements (AND, OR, NOT).

The results were so good that it was decided to move from a batch retrieval system to an interactive one: *Automatic Direct Access to Information with the On-line UDC System* (AUDACIOUS). This online system allowed users to type in natural-language terms that were translated to an appropriate UDC notation. The UDC system was then opened for browsing and users were able to link selected numbers with Boolean logical operators and launch a search. This was the first online interactive retrieval system based on a widely used library classification.

The final report (Project Report number 6) contained a valuable contribution in the form of a failure analysis aimed at system designers and managers that were planning to use the UDC. Freeman & Atherton (1968, 1969) experienced difficulties in defining algorithms for handling UDC symbols as sometimes the same symbol meant different things in different contexts, e.g. a decimal point, colon or = (equals) sign changes its meaning when preceded or followed by other symbols. Also, the software application used could not automate complex numbers built with / (extension) and + (additions) so this had to be handled by Boolean operators.

UDC automation by Caless occurred almost simultaneously with AUDACIOUS and is the second most important project with regard to the mechanisation of the UDC. The project was carried out by Thomas W. Caless from George Washington University. Caless worked for the American Geological Institute and was employed to produce the *Bulletin of Abstracts of Geological Literature Excluding North America* in 1967. Based on a subject analysis of the Institute's Earth Sciences collection, he devised a set of matrices for the subject analysis of Seismology defining the following facets: whole thing, kinds, parts, materials, processes, properties, operations and agents. This technique was then used with the UDC to provide both flexibility and consistency in indexing. Assisted by Perreault who classified several hundred documents, Caless devised a procedure for subject analysis, citation order, and a specific set of relationships that were handed to UDC trained staff. His solution for searching UDC was based on UDC notation coding that enabled the management of facets and citation order as well as the management of relations between UDC numbers.

Caless' approach allowed search access to every element of pre-combined numbers. Each combination, such as a main number and special auxiliary, or main number and main auxiliary, was coded in the system, which helped mechanisation. The result contained a described record format managed by the database management system that provided Boolean search functionality as well as *greater than*, *equal to*, and *less than* options (Caless, 1969; Foskett, 1973).

Up to 1980 the UDC was the most advanced library classification system used in an online environment. A number of studies and experiments reported that UDC could be effectively used as a term system as well as a classification (Freeman & Atherton, 1969; Stueart, 1971; Ścibor, 1976; Černyj, 1976; Hindson, 1976; Öhman & Olivecrona, 1976; Rigby, 1981). Hence, it was not surprising that computer applications on UDC information retrieval up to 1980s involved mapping UDC to thesauri, descriptor lists and indexes.[8]

In some cases, UDC was given a central rôle between different indexes in an information system (see Rigby, 1971, 1974, 1981; Gilchrist, 1972). For instance, the

[8] The rôle of a subject-alphabetical index to classification in online IR was, for instance, brought to the forefront by the Cranfield I project: "... it could be argued that the unusually good performance of the Universal Decimal Classification in Cranfield I owed much to the unusually thorough A/Z index made for it..."(Cleverdon & Mills, 1963: 117).

UNISIST project on international information network considered the UDC as a serious candidate for a 'switching language' (Figure 1.1). *The International Symposium: UDC in Relation to other Indexing Languages* in 1971 in Herceg Novi (Yugoslavia) was particularly focused on this specific rôle for the UDC. At this conference there were numerous examples of analysing the scheme as a possible interdisciplinary reference standard for thesauri, or for creating a list of concordances with thesauri or descriptor lists (Proceedings of the International Symposium UDC in Relation to other Indexing Languages, 1976, Rigby, 1981; Santoro, 1996).[9]

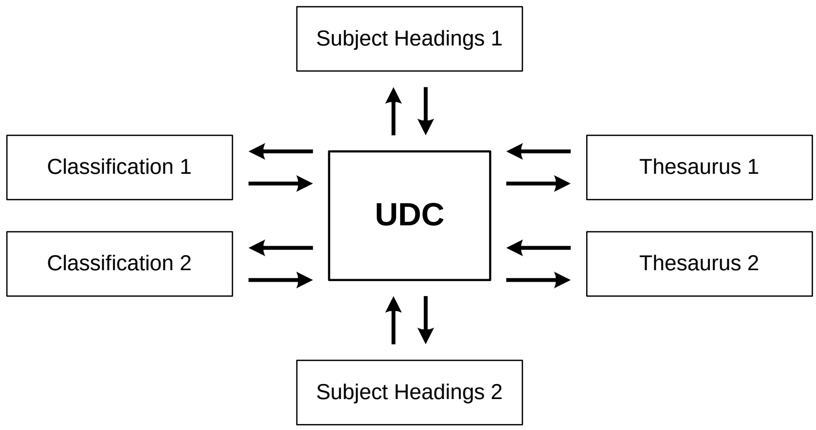


Figure 1.1 UDC as a 'switching language'

Eventually, the idea of using the UDC as an intermediate language was abandoned in favour of creating new classification systems[10] as it was estimated that there would be an overwhelming amount of restructuring needed in order to make it suitable for such a rôle (Dahlberg, 1976).

1.2.2.2 UDC in IR from 1980 to 1993

In 1980 Rigby predicted a growth of SDI services using UDC, the use of UDC in library networks and multilingual terminology banks based on the UDC for the period following the automation of the UDC schedules maintained in The Hague. The period after Rigby's bibliography on automation did not entirely follow his

[9] Good examples at the time were concordances between UDC and EJC /TEST (Thesaurus of Engineering and Scientific Terms) described by Wellisch (1976) and Kara Marosi's (1976) study on the relationship between UDC and the Euratom thesaurus.
[10] Namely, Standard Reference Code or Standard Roof Classification (SRC) and later Broad System of Ordering (BSO).

predictions and did not fulfil his optimistic expectations. UDC automation did not progress with the rapid pace its history had suggested, and the UDC itself shared the unfortunate destiny with other classifications of being an unwanted and superfluous tool within library systems. But what is most regrettable is that continuing work on the UDC did not make full use of the research on automation already undertaken before the 1980s. Large scientific and research institutes that were interested in the UDC gradually ceased to develop their own bibliographic services and turned to cheaper, commercial vendor bibliographic services.

For the majority of librarians this period was denoted by intense activity in library automation with little or no attention paid to the automation of classification. Outside the library domain, research into information retrieval was dominated by the development of more sophisticated retrieval techniques, automatic indexing, ranking and probabilistic methods. Evaluation of the efficiency of information systems established an unfounded belief that what cannot be precisely measured is not worth using. Experiments on automatic indexing and expert systems attracted more attention than the management of existing controlled vocabularies. At that time, as stated by A. Gilchrist (1992) the UDC Revision Committees were dormant.

The end of the eighties and the beginning of the nineties can be depicted as an unfavourable age for all subject indexing languages as they were suppressed in library systems and consequently disappeared from library education in the majority of American and British library schools (Gorman, 1993; Buckland, 1996; Weinberg, 1996). Although the first library online catalogue that allowed the search of UDC numbers was inaugurated as early as 1974 in the *Bibliothèque des Halles* in Paris (Rigby, 1981), with more than one million titles searchable by UDC, this was rather exceptional in the domain of library automation.

The command driven menus of library OPACs were not a good environment for a classification that relies on the visualisation of subject relationships. Classification was represented in catalogues with notation and symbols in a very user un-friendly way. Keywords in title and author searches were widely accepted as sufficient to satisfy user's needs in the majority of vendor library systems. For the UDC, implemented as it was in in-house library systems, the situation was much better, as illustrated in Buxton's research (1990), which provided an overview of some successful solutions in library OPACs and bibliographic databases.

Following this 'dark' period, the greatest change that influenced the perception of classification came with the introduction of friendly graphical user interface (GUIs) and most importantly, with the Internet. Incidentally, the first classification to be used for automatic classification of Internet resources was the UDC in the *Nordic Wide Area Information Server*/*World Wide Web Project* (WAIS/WWW) 1993-1994. The year 1993 could be considered as the threshold of a new era of automation of the UDC as this is both the year of the automation of the UDC schedules and the year of the first application of the UDC on the Internet. This research will reflect on the developments and changes in automation of classification after 1993 in Chapters Four, Five and Six.

1.2.2.3 Automation of the classification process

The link between a library classification and automatic text processing has to be viewed with a certain degree of circumspection. The automation of the classification process may refer to different kinds of tools and procedures.

The first group of automatic classification tools is often termed (sometimes incorrectly) *expert systems,* and these are designed to assist and automate the classification of documents according to a given library classification system. The second type of system consists of programs designed to hold and handle a specific classification system and its rules, to help automate class number assignment without processing documents. And, finally, the third type of system which is most pertinent for information retrieval and resource discovery on the Internet is a group of advanced tools for the automatic indexing of documents for specific classification systems. This last group of tools is based on the application of natural language processing (NLP) methods and uses bibliographic classification either to control terms or to perform subject indexing for information retrieval. These tools are likely to deploy AI methods and applications for knowledge representation (e.g. rule-based models, frame-based models, semantic nets and neural networks).

A typical representative of the first group, that is the tools for automating a classification process for a particular classification system, is *ShelfPro.* This is a program for the automation of class number assignment which is part of an expert shelflisting system designed for DDC (Markey Drabenstott, Riester & Dede, 1992). This is an example of a library tool, based on separate diagnostic problems typical for a classification system, which could lead to the automatic assignment of class numbers. The process includes the analysis of a book description and its relationship to other books in a collection. The librarian does the initial input guided by the program's questions and set of pre-programmed steps and rules. Rule-based expert systems can easily be devised to accomplish the task automatically and can be envisaged for any classification system. There are, however, very few reports on classification expert systems[11] in the literature and none were found for the UDC during the course of this research.

The next group could be considered a collection of tools designed to support the automatic management of classification systems. These tools are created to assist human indexing using certain classification without processing the documents to be classified. Schneider (1979) reported on *Autoclass*[12] which is a series of computer programs designed to handle a wide variety of hierarchical classifications and to unify multiple indexing languages by expressing them through the structure of a *common central classification (CCC).*

Ishikawa & Nakamura (1990) and Ishikawa (1994), reported on UDC-AUTCS, a system for the automatic combination of UDC numbers that can be used for machine-assisted indexing. Srimurugan (2001) reported on ongoing research in designing UNIDEC, a software package to support classification using an abridged UDC edition. The system is supposed to be a *dialog generation and management system* (DGMS) and a tool for machine-assisted indexing.

An interesting undertaking in this area was the project *Knowledge-Based System for automatic CROSS-referencing of classification systems - KBS-CROSS,* and it was started at the *KBS-media Lab* in Lund, Sweden.[13] This was a one year project (1994-1995) for which the outcome was unclear. The project announced the application of neural network technology to improve the learning behaviour of software supporting navigation, browsing and resource organization. The goal was to

---

[11] The expert system most often mentioned in subject indexing literature happens to be unrelated to classification: MedIndEx, developed at the National Library of Medicine (Lancaster & Sandore, 1997). Other examples are SIMPR (Structured Information Management Processing and Retrieval), CLARIT (Computational Linguistic Approaches to the Retrieval and Indexing of Text) (Schuegraf, 1997) or Indexicon (Browne, 1996).

[12] Not to be confused with AutoClass which is a knowledge-based system by VisionShape that uses neural network based technology and is designed for image processing.

[13] This project was a take up of another project by KBS-Media Lab at the Lund University: *Knowledge-Based System for automatic CLASSification of building texts - KBS-CLASS* (1993-1994), http://it.civil.auc.dk/it/delphi/KBS/projects/kbsclass.html.

produce a tool that cross-references the LCC with the UDC within the subject area of building and architecture. The practical application was supposed to be machine-assisted cataloguing and classification at the Lund University Library which would assign UDC notations based on LCC and Library of Congress Subject Headings (LCSH) data already associated with a document and vice versa. It was also planned to link this tool to an OPAC search. The test-bed for the project was the Swedish union catalogue LIBRIS (The KBS-Cross Project, 1995).

The third group of tools performs automatic document indexing using bibliographic classification for the purpose of IR. Automatically extracted terms (keywords) can be used for information retrieval or can be mapped to some existing controlled vocabulary such as a classification system, resulting in a fully automated classification process (Croft, 1989: 90).[14]

One example of such a system is a program called *Viswamitra* which applies an automatic indexing system using Colon Classification (CC), reported by Panigrahi & Prasad, 1997. The system analyses expressive document titles using a natural language parser and the classification process is performed by matching terms to CC classes and applying a set of predefined rules. Another automatic indexing system based on faceted classification postulates that produces synthetic strings of verbal indexing terms is the *Postulate Permuted Subject Indexing* (POPSI). The knowledge representation model chosen to support inference rules for syntax synthesis in POPSI is the frame based model (Aptagiri, Gopinath & Prasad, 1995).[15]

With regard to the UDC, the only examples in this category include the

---

[14] This method of automatic indexing using library classification should not be confused with automatic classification proper, which is actually an automatic clustering method and is not of direct relevance to this research. Automatic classification or clustering is a separate field of AI that has been successfully used in advanced information retrieval systems (cluster-based models) since the late 1950s and early 1960s. Automated classification proper is the grouping of documents based on an analysis of document patterns and their relationships (Borko, 1979). The process of finding the right data, hidden knowledge and its correlations in large data collections is addressed through different data mining techniques (pattern finding).

[15] Similar to POPSI is the Prometheus system based on the Ranganathan school of classification and Bhattacharyya's theory of "deep structure of subject indexing languages". Prometheus parses expressive titles and extracts noun phrases within documents which are then processed through a knowledge representation model to generate meaningful strings (Prasad, 1996). Also Schuegraf (1997) reported on experimental automatic indexing systems in which the terms extracted from the text of document abstracts are matched to categories from the INSPEC classification.

automatic indexing and classification of Internet resources, which is going to be described in more detail in Chapter Three.

A number of studies were undertaken in the 1970s and 1980s calculating the efficiency (recall and precision, relevance) of automatic indexing compared to human indexing. Whenever complex and advanced automatic indexing was used the automatic system seemed to perform as well as or better than human indexing (Rijsbergen, 1979; Ellis, 1996). The most interesting point stressed by Croft (1989) was that although both automatic and human indexing may produce equally relevant results, the choice of indexing terms is often different and rarely overlaps significantly.

1.3 The rôle of classification in IR

The efficiency of an IR system is usually defined as its capability of retrieving relevant, and suppressing non-relevant information. Svenonius (1983, 1995) summarised the criteria for evaluating information retrieval using controlled languages as the following: neutrality, specificity, browsability, ability to support precision and ability to support recall.[16]

For a long time, the research focus was on whether *alphabetical* or *classificatory* indexing languages and whether *pre-coordinated* or *post-coordinated* indexing languages are equally efficient in IR (Rowley, 1994; Svenonius, 1995, 2000; Bodoff & Kambil, 1998; Miller & Teitelbaum, 2002). For improving precision, it is acknowledged that controlled indexing languages have advantages over free text searching, while pre-coordinated indexing languages have advantages over post-coordinated indexing languages (Fugmann, 1983; Rowley, 1994; Svenonius, 1992). The rôle of classification as a pre-coordinated indexing language is then recognized as important for both improving recall and precision in IR

---

[16] Precision is defined as the degree to which systems retrieve only relevant documents, and recall is the system's ability to find all relevant documents. The precision/recall measurement and quantification allows one to establish the impact of some factors such as the precision of indexing or the number of index terms on the efficiency of a system. Both measures, however, depend on the notion of relevance being a simple and objective value which was often disputed. The concept of relevance was later re-defined to be complex, multi-layered, relational in nature and to be part of an interactive and dynamic communicative exchange between user and system (Saracevic, 1996).

(Svenonius, 1983, 1995; Fugmann, 1990, 1993; Markey Drabenstott, 1996a). The potential of an indexing language to influence precision and recall has important practical value when choosing an indexing system and defining an indexing policy (Soergel, 1994; Svenonius, 2004; Fugmann, 1993).

Heterogeneity of information resources on the Internet: text and non-text documents; full text and surrogates; different languages; different scripts; diversity of users and their information seeking habits - had a great impact on reducing the competitiveness of indexing methods. Instead of choosing between a human or an automatic approach, research in IR has become open to the integration of advanced IR methods, automatic indexing and traditional tools, with or without human indexing (Sparck Jones, 1995; Svenonius, 1986; Milstead, 1994).

1.3.1 Interactive IR

Throughout the early phase of online bibliographic database services (i.e. before CD-ROMs and later the Internet), a search strategy was performed by subject specialists. Dis-intermediation of information retrieval, free browsing and navigation through a collection evolved through better interface support and improved functionality (Hildreth, 1991, 1995, 1995a). When users started to interact with a system directly, attention was shifted from search results to the searching process itself. The Graphical User Interface (GUI), introduced by windows-based operating systems and later on, Web technology, greatly influenced information seeking and laid solid foundations for interactive information retrieval.

From the first conference in 1992, the Text Retrieval Conferences (TREC) follow different tracks initiating international collaborative research and in-depth studies on different issues in IR, which were then reported at subsequent conferences (Sparck Jones, 1995). The *interactive track* of the TREC conferences has been especially important for understanding and supporting combined information seeking strategies which could influence interest in controlled vocabularies. A person's interaction with information has been proved to be central to IR and many user strategies depend on opportunistic interaction which can occur through browsing. Subject searching is defined as both highly complex and highly interactive (Belkin, 1998).

Although no comprehensive interactive IR study exists involving classification, the research reported by Spink & Saracevic (1997) on thesaurus use in term selection confirms that when a user is not focused on one particular search term, retrieval tends to be more interactive. The authors emphasised the value of a controlled vocabulary in supporting the user in defining their information problem and search expansion. Similar findings were later confirmed by Shiri, Revie & Chowdhury (2002) and Binding & Tudhope (2004). These studies stressed the importance of semantic relationships between terms and these are part of the basic structure of both a thesaurus and a classification system.

1.3.2 Subject browsing

The larger the collection, the greater the need for precise queries and users are more likely to need assistance in formulating their information problems. In an exploratory search, when a user does not know precisely what to look for, IR systems need to interact and provide some kind of clarification and disambiguation support, usually through offering term choice or topical browsing. Subject browsing is an inseparable component of information retrieval behaviour and has become even more important with the growth of electronic collections (Marchionini, 1995; Ellis 1996).

User behaviour in the process of browsing was the subject of much research in the nineties, as information retrieval became increasingly user-orientated (Marchionini, 1995; Belkin, 1998). Observation of user behaviour led to the re-evaluation of the classic information retrieval model. The concepts of *human information behaviour* (HIB), *cognitive load*, *cognitive laziness*, and the *anomalous state of knowledge* (ASK) are representative of the IR research focus in relation to this problem (Lee Pao, 1989; Chowdhury, 2003; Chu, 2003).

According to Marchionini (1995) an analytical search, i.e. one using terms and words, is highly planned, deterministic, goal driven, formal and discrete, whereas a browsing strategy tends to be opportunistic, data driven, heuristic, informal and continuous. In searching, the emphasis is entirely on the users and their capability to define their information problem in one or more terms. In browsing, the emphasis is on the system, which has to interact with users (i.e. visualisation of semantic relationships, verbal expression instead symbols, navigation paths highlighted and enabled through hypertextual linking). Users engage in subject browsing to gain an overview, monitor a subject area, shift/share cognitive load, clarify an information problem, develop a formal strategy, discover/learn, or respond to environmental invitations.[17]

Bates (1989), for instance, introduced the term *berrypicking technique* to describe a common and realistic information-seeking model whereby query formulation is not one step from search query to search results but a series of evolving moves from searching to browsing and back, based on system feedback from which users pick and gather information on a subject as they go. The searching not only evolves but uses a combination of searching and browsing techniques.

Apart from its obvious advantages, browsing has a number of limitations that should be taken into consideration. According to Marchionini (1995) these are: high attention demand, inefficiency in well-defined retrieval, possible distraction, possible information overload, influence by various biases and cognitive inertia, and the fact that present systems are not designed to assist browsing, which diminishes return.

Prior to the electronic information environment, subject browsing was almost exclusively related to library shelf browsing and to the systematic arrangement of library collections or a classified/systematic catalogue. In an IR system, this corresponds to any situation where users are looking for resources but may have no knowledge of what these resources may be, what they are called and by whom they are authored. The universe of subjects, as has traditionally been organized by library classifications, is therefore even more important as it offers a much needed map for subject navigation.[18]

The exploratory value of browsing was often connected to the accidental and

---

[17] Marchionini (1995) discriminates between four different browsing strategies: scanning, observing, navigating and monitoring. Scanning is a basic recognition function and is typical of someone who knows what he/she is looking for. Monitoring is similar to scanning but does not depend on the organization of the environment. It is typical for establishing creative connections. Observation is typical for users who find themselves in a promising environment and is characteristic of opportunistic and casual browsing. Navigation is the strategy when a user interacts with the environment and actively determines direction. Navigation is often considered to be a synonym for browsing in general but is actually particular to an organized, structured environment.

[18] Browsing as a strategy is found to be more common among humanities than among science and engineering scholars, as the former are more interested in general and background reading in the field, and text and reading is the primary objective of their research. In addition, humanities deal with less rigorous vocabulary which makes searching less reliable (Langridge, 1976; Shoham, 2000).

unexpected discovery of valuable information. Unlimited and unpredictable combinations of basic concepts facilitates coincidence in information discovery, serendipity and thus discovery of new knowledge. Research on literature-based discovery shows the important rôle of IR systems in facilitating unexpected scientific discoveries. Swanson (1986, 1987, 2001) did extensive research into the possibility that information in one speciality might be of value in another without anyone being aware of the fact. He published a number of articles on coincidental discovery through the process of bibliographic database searching and browsing.

Roy Davies (1989) pointed out the importance of a general classification in providing a horizontal view across disciplinary boundaries that may reveal gaps in existing knowledge and strengthen a controversial theory or lead to the emergence of a completely new field of study. He singled out faceted classifications that contain the mechanism of concept combinations in complex syntactical expressions (self-perpetuating classifications) and the fact that their browsing and post-coordinated searching may play a greater rôle in serendipitous knowledge discovery.

1.3.3 Change in use of classification

During the early development period of online IR systems, it was anticipated that classification would be used for searching in a very primitive way - expecting users to form search expressions using classification numbers. Many arguments against classification were based on user studies in which users were presented with classification numbers to search, with no subject-alphabetical index to the classification, poorly supported browsing functions and a command/menu driven user interface. An extremely low ratio of users searching classifications in OPACs, for instance, was then used as an argument against classification.

Gradually, this approach was abandoned and the use of classification in IR evolved. In modern IR systems classification and natural language terms tend to be used in a complementary fashion for post-coordinate searching, while the main rôle of classification is that of search expansion and subject browsing (Fugmann, 1983; Svenonius, 1995; Rowley, 1994; Bodoff & Kambil, 1998).

Classification systems remain important for indexing and information retrieval because of their conceptual transparency, language independence and, when used with other indexing languages, because of their rôle as a 'categorial' guide. The main rôle of classification in IR is to support subject browsing across remote and unrelated subject areas, as this particular function cannot be supported by alphabetical indexing languages which are generally accepted as more user-friendly. It is typical for a modern IR system to use a classification scheme in combination with other alphabetical indexing languages and to automate both the indexing and retrieval process (Aboud et al., 1993).

An interesting summary of the subject query types that may help decide the use of classification in an IR system is given by Gödert (1996) in his research on improving subject searching in OPACs:

- simple subject (concept) query
- formalised queries against established relationships (editor, edition, publishing year)
- queries on subjects containing individual names that need disambiguation in the context of subject arrangement
- queries on simple subjects with relationships to place or time
- queries with complex *a posteriori* relationships between concepts e.g. *'Laying eggs of blue-tits while feeding their chicks'* or *'punishment of threatening behaviour by ignoring*'
- queries in which apart from the subject itself one expects all sub-subjects to be included e.g. *'literature on psychology'* or *'literature on the kingfisher'*
- queries on a subject and all sub-subjects in a field with *a posteriori* relationships with another subject and its sub-subjects e.g. '*sacral buildings on the Lower Rhine*' (all sacral buildings in all places on the Lower Rhine)
- queries in which an exclusion of the subject may occur with another subject, if required

If fully implemented, classification can obviously support any of these subject queries. The requirements of IR arising from these query types would include both searching and browsing a classification and this may apply, for instance, to both classification numbers and their verbal equivalent (words). In a modern information system that bridges distributed collections, the lower specificity of classification indexing terms, its use in broadening and narrowing searches and browsing, have great potential for reducing information loss (Markey Drabenstott, 1996a).

Classification is, however, a highly formalised indexing language meant to be used in its 'raw' form only by classifiers and even then, supported by appropriate classification tools. End-users should browse and search classification via natural language terms and classification can be used to hold a semantic structure in place and link, maintain and control alternative subject access, especially in multilingual and networked environments.

As UDC is a synthetic classification, it demands a two stage adaptation: first the adaptation of the original classification schedules to a classification scheme with complex, pre-coordinated numbers that are used in indexing. The second is adapting the classification scheme for end-users, by adding and expanding the subject alphabetical index to the classification and designing a user-friendly interface to enable the browsing of hierarchies and 'see also' links, and easing the transition from searching to browsing. It is obvious that the use of classification will depend on the way these two steps are facilitated and controlled by the information system. To achieve this, classification data ought to be held and maintained as an external authority control tool (Markey Drabenstott, 1992; Guenther, 1992; Woods, 1996; Taylor A.G., 1999).

1.4 Knowledge organization on the Internet

The use of library classifications on the Internet is connected to the growth of academic and research information infrastructures consisting of commercial and non-commercial information resources of interest to the academic community and national research networks.

The Internet is now moving into a new era (Internet 2) with a new World Wide Web architecture known as the *Semantic Web*. In this environment, the Web will be gradually populated with content supporting a machine-understandable description of meaning that will enable the automation of content organization and its retrieval (Berners-Lee, Hendler & Lassila, 2001; Lassila & McGuinness, 2001). As implied by its name, this kind of network will assume a higher level of connectivity which will be based on resource content and meaning, while the information organization will predominantly be automatic i.e. based on *machine to machine* (m2m) information services.

Information organization on the Internet has been determined by the particular nature of its information content and the way it is accessed and used in the networked environment. Information resources can be static or dynamic and may exist for a short period of time, or be persistent. The machine-driven, rapidly growing, distributed collection of resources on the Internet is often amorphous, ill-defined, is not self-contained, is unstable and volatile (Chan, 2000). This defines the difference between information seeking in a traditional, closed IR system and on the Internet. The Internet deals with different types of information resources and different information seeking services, which promote more interactive user information behaviour.

Information objects/resources vs. documents. The document in a traditional bibliographic IR system is mainly textual and the entire bibliographic apparatus is built around the coexistence of 'work, publication, copy' and the relationships that may be established between them. An Internet *information resource* may include services, individuals, organizations, buildings, datasets, software, events and collections. It may be, for instance, a database or a collection of databases and the services built upon them, or a single record in some database or, indeed, metadata about a database. Furthermore, resources may be mutable, dynamic and adaptive to access permissions. Not only do they change the information space by contributing to its diversity but even more importantly, they influence the balance by making the Internet predominantly non-textual (Dempsey & Heery, 1998).

It has been more appropriate to refer to an electronic resource on the Internet as an *information object* which stands for 'a digital item or group of items, regardless of the type or format, that a computer can address or manipulate as a single object' (Introduction to metadata, 1998).

Resource discovery vs. information retrieval. Users' information seeking behaviour on the Internet consists of a whole series of actions: look for, retrieve, save, edit, recall or print. This combination of 'search, discovery and delivery' is called 'resource discovery'. Dempsey (1994) summarised the fundamental aspects of resource discovery as: global in scope, having a distributed nature, dealing with non-structured resources, having no single protocol standard, being scalable and encouraging serendipity. He also expected these characteristics to evolve with Web technology.

A closed IR system is often designed for a specific known type of document, to be used by a certain audience. IR is usually performed on a limited number of documents or their surrogates that are represented in a structured and familiar format. Hence, IR system efficiency often depends on training the users to use the system or on users' knowledge of resource access points (e.g. title of the book, author, subject). Internet users (programs or people), however, are not supposed to be aware of any resource organization or management tools.

Open access and cross-domain searching. Internet resources reside on different systems, they vary in format and size and are made available through different access standards and rules. The ultimate goal of many information providers and services on the Internet is that the end-user is unaware of the actual location of the resource or the complexity of the information architecture that assists in its discovery and retrieval. There is a general tendency to provide seamless access to resources distributed over many collections and shared between different information sectors and domains.

It is logical that new standards for resource description (metadata) and resource exchange are constantly being developed to enhance connectivity and information exchange and these are combined with constantly emerging methods in IR. Interoperability has become central to knowledge and information exchange. Full text indexing and searching of the Internet resources that were present at the beginning, grew to become the foundation on which more sophisticated methods were built. Nowadays, every provider of so called 'quality information'[19] uses metadata and the content indexing model to meet the requirements for cross-domain interoperability.

The importance of the subject approach to information. The Internet's information space is limited only by the size of the global network. The variety of forms and versions the resource may have and the fuzzy notion of its provenance, location and availability, make traditional bibliographic access points such as title or author less important than the subject content of the resource.

Searching the Internet, one might expect to find the same document-like object appearing several times in different contexts: sometimes standalone, sometimes as part of a bigger document. The title, author, the date when it is made available, as well as its format may vary. The only tangible characteristics of the

---

[19] 'Quality information' is the expression adopted for an information resource for which a source/publisher can be established and which can be connected with some known authority/institution in the field. More recently the term 'trusted' is also used with a similar meaning.

resource the user can define would probably be related to its subject content (i.e. looking for something about ...)

To ensure subject presentation, display and browsing across domains of formats and languages in a global space, it is clearly necessary to accompany resources with some standardized, machine-readable and human/machine-understandable description that will also contain information on what the resource is about. The idea of the subject may be related to both commonly accepted and communicated public knowledge and personal subjective experience and understanding. The 'commonly understood', when it means *educational and scientific consensus*, is intrinsic to classificatory indexing languages which tend to assume an 'objective view' of knowledge offering an independent framework in which communication between a document’s content and a user's knowledge can be established.

Multilingual environment. Full text indexing, or even more sophisticated automatic indexing techniques, cannot bridge differences in scripts and languages. Even variants of the same languages, jargon, slang or style impose serious obstacles to precise indexing. It is recognized that as non-English websites expand, different types of indexing strategies will be necessary (Arnold, 1999). The ultimate goal of multilingual information retrieval systems is to offer users the opportunity to write a query in a given language and retrieve a merged and ranked set of documents that match the query in any language. This approach requires advanced technology and research and is usually associated with higher developmental costs, and often uncertain efficiency.

Alternatively, an information service may provide access to multilingual resources based on a resource description (metadata) rather than on searching the documents themselves. In this scenario, an indexing language may be deployed to describe the content, and it is only this indexing tool that needs to be searchable in different languages. Metadata carrying content information is equally important in the retrieval of images (still and moving), audio resources, datasets, collections and events.

The importance of metadata standards in resource discovery on the Internet is growing with the expansion of services and is considered as fundamental in the development of the Semantic Web. Its nature, scope and rôle in carrying subject information is paramount for this research. Dempsey & Heery (1998), however, relate the scope of information management on the Internet to the true scope of what an electronic resource may be from a simple document-like object, server, a site, or file on a server. They argue against the naive, or immature, as they say, view of a generic metadata format that may serve to describe all kinds of electronic resources on the Internet. According to them, it is not possible to impose 'one single view which would guide the design and development of metadata formats across domains'.

1.4.1 Network information discovery and retrieval (NIDR) metadata

Metadata describes the attributes of a resource and supports location, discovery, documentation, evaluation, selection and other activities which may be carried out by humans or by programs (Dempsey, 1996). In spite of the many similarities between metadata and a bibliographic description, the new concept of metadata was adopted to avoid misinterpretation and to underline a difference that is crucial for metadata understanding and application (Vellucci, 1998). Resource creation, publication on the Web, updating, resource life span, metadata production, discovery and access - all occur in a very condensed time span, without mechanisms to control or influence the process. The sheer number of resources makes clerical metadata production and indexing impossible (Dempsey, 1994; Miller, 1996, Vellucci, 1999), and metadata as machine-understandable information about Web resources has become fundamental for the development of the Semantic Web (Berners-Lee, Hendler & Lassila, 2001).

NIDR metadata can be simplistic or rich and structured and can be found in different forms and places: resource *Hypertext Markup Language* (HTML) source code, document headers (*Text Encoding Initiative* - TEI), image files (*Portable Network Graphics* - PNG, *Tagged Image File Form* - TIF), e-mail headers, records of organizations (governments, industry), *Geographical Information Systems* - GIS data, DMBS dictionaries, and in server/service descriptive data (Z39.50, *Common Object Request Broker Architecture* - CORBA) (Dempsey & Heery, 1998). It can be **embedded** in the resource, created and delivered as part of it, or may exist as **standalone** independent descriptions, managed separately from the resource itself. One ought also to distinguish the following types of metadata: *administrative* (known as meta-metadata), *descriptive* (resource description i.e. finding aids), *preservation* (physical condition of the resource), *technical* (system-oriented data, software or file related information), **access** (use and access control and conditions) (Gilliland-

Swetland, 1998). This is why metadata can be *static* such as, for instance, descriptive metadata that tends to stay the same once produced, and *dynamic* such as preservation metadata that changes during the lifetime of a resource.

At present, there are three models of metadata in use: the **embedded model** with metadata produced by the creators of the resource as an integral part, the **third party metadata model** with standalone metadata produced and managed by the agencies and the **view filter model** where independent brokering services harvest metadata, map it and provide cross metadata, cross-sectorial and cross-collection searching (Weibel, 1997). More and more metadata on the Web is created incrementally, and this task is performed by different agents (Bearman et al., 1999). This is why the model of metadata embedded in a resource is gradually giving way to standalone metadata which is easier to amend independently of the resource.

Different information needs and different functional requirements created a diversity of metadata formats in different communities. Pre-Internet metadata standards such as Machine Readable Cataloging (MARC), Encoded Archival Description (EAD), and TEI, are being adapted and improved to support new functions. In addition to these, some sectors such as government, education, publishing and commerce have developed their own metadata standards when a substantial part of their activities moved to the new communication environment.

Services that bridge many distributed collections, such as portals, are commonplace on the Web. Information discovery in this context consists of searching across different metadata structures and interoperability on the level of metadata encoding[20] has become very important. Cross-metadata resource discovery requires the technological solutions (middleware) and architecture that support the interchange of metadata packages between different discrete metadata repositories.

---

[20] XML is generally accepted as the norm for metadata encoding. The reason for this is the simple textual and self-explanatory nature of XML which gives the freedom to metadata developers to create an encoding syntax of their own while declaring the processing rules and grammar in the Document Type Definition - DTD or XML Schema. Any XML metadata schema can be given an identity and declared as a 'namespace' with a specified identifier (URI) and each metadata element as a constituent of the given 'namespace'. The existence of XML and the URI have enabled the development of RDF which can serve as a complex machine-understandable metadata 'container'. RDF supports encoding, exchange and re-use of structured metadata and the combination of different metadata structures within one single record that can have links to an external source of reference.

Examples of communication protocols solving this problem are Z39.50, which was developed by the bibliographic community (Denenberg, 1996) and *Open Archives Initiative* (OAI) which has been developed by Internet community at the end of the 1990s.[21] In addition, metadata crosswalks are being provided either by mapping different complex schemes to one another or by mapping from a domain specific scheme to a general simple metadata model, such as the *Dublin Core Metadata Element Set* (DCMES) (Day, 1997; Cromwell-Kessler, 1998; Moen, 2001).

One of the most important prerequisites for a **subject search** across different metadata standards is the possibility of 'understanding' and mapping the semantics and syntax of different vocabularies used to describe a subject. This problem is addressed in the metadata community by ensuring that:

i) subject metadata is properly encoded and identified
ii) subject metadata uses standardized and widely shared subject indexing languages

It is for this reason that classification schemes such as DDC, UDC and LCC were identified for possible use in subject description in a number of metadata standards. Because of the fact that heterogeneous collections using different subject indexing languages may be searched through the same service, mapping between different subject indexing languages has become a necessity (Hiom, 1998; Buckland, 1999; Doerr, 2001; Vizine-Goetz et al., 2004).

The choice of classification for a given metadata implementation is always related to the intention of the service to provide subject navigation and is part of the implementer's commitment to more interactive retrieval. Hence, a classification will be used only if it allows easy system implementation of functions for both hierarchical and lateral browsing and a transition from word searching to navigation. Logically, classification schemes which come with machine-readable data to support these functions have greater advantages. Metadata as a carrier of subject vocabularies such as classification will be addressed in greater detail in Chapter Three.

---

[21] An earlier development in this area was Warwick Framework (Dempsey & Weibel, 1996; Lagoze, 1996).

1.4.2 Information discovery

The Internet information space consists of organized and unorganized digital resources and manual, machine-assisted or fully automatic services created to assist in resource discovery and knowledge organization. The rapid growth of the Internet was followed by an equally rapid proliferation of activities, initiatives and standards targeted at its organization.

There have been a number of studies focused on Internet information service evaluation and categorization (Koch &Vizine-Goetz, 1998; Anagnostelis, Cooke & McNab, 1997). Based on the rôle information services fulfil, Koch suggests the following two categories of information services: **general search services** and **subject gateways** (Koch, 2000)**.** These two categories of service are being developed and are evolving to coexist, each fulfilling its particular function.[22]

1.4.2.1 General search services

General search services, also known as *search engines*, being both numerous and technologically diverse, are all based on free-text indexing and automatic Web page harvesting. They are not usually designed for a specific audience or with a specific purpose in mind. Although they are not selective with respect to the pages they harvest and strive towards wide Internet domain inclusion, they can vary greatly in quality and coverage. As a rule of thumb, these services do not harvest terms from resource-embedded metadata as this proved unreliable and often contained false or misleading information.

General search services consist of the following architectural components: (*a)* 'spiders'- programs designed to access Web pages to build an index of their content, (*b)* index database - containing strings of characters (words) and *Uniform Resource Locators* (URL) of pages where this string occurs and *(c)* interrogation software. It

[22] The division on 'visible' and the 'invisible' Web is an important distinction (Green, 1999; Lawrence & Giles, 1998). The visible Web consists of static pages created manually by Web designers that are posted on a server and are usually updated manually, indexed and searchable through search services, subject gateways or launchpads. Conversely, the 'invisible' Web consists of dynamic pages created by computer programs on the basis of information filled in by users. The invisible Web is made up of information stored in databases and is usually not indexed by general search services, although some of the more advanced can provide access to a limited number of databases (Sherman, 1999). The number of dynamic pages is growing faster than the static part of the Web.

emerged that the coverage of any one service was limited and in 1998 it was reported that no single service indexes more than approximately one third of the 'indexable Web' (Lawrence & Giles, 1998). The overlap between some large services was reported to be as small as 1.4%, which led to the creation of *meta-search engines* (Bharat & Broder, 1998).

Each general search service has a different index-refresh frequency, uses differently tailored methods of interrogating its database and different algorithms for ranking results. These mechanisms have progressed through several phases of development to include different refinement techniques corresponding to different generations of services (Green, 1999; Arnold 1999; Bharat & Broder, 1998; Bradley, 1999). The most successful development is search engines that perform hit ranking using link-based analysis, such as Google. In the search results of these services, Web pages containing a search term are ranked by importance i.e. according to how many other Web pages are reporting them (Green, 1999).

General search services perform the function for which they are designed quite well, as they will find exact matches to desired search expressions from millions of Web pages on the Internet. If one knows how to limit the recall logically and then knows how to refine the results, search engines provide good support for routine checking and challenging serendipity. They are especially useful for checking occurrences of names of persons, objects and products in cases where an exact matching will give relevant results. While their greatest advantage is their coverage, at the same time their 'general' nature makes them unsuitable when it come to providing quality and trusted subject information.

1.4.2.2 Subject gateways

Subject gateways[23] i.e. quality information services, emerged as the response to the challenge of resource discovery in the mid-nineties. Smaller in scope and more discriminating, these services focused on the selection, description and subject presentation of resources. (Koch, 2000; Vizine-Goetz, 1998a). They provided a

[23] Synonyms to *subject gateway* are 'subject-based information gateway' and 'subject service' sometimes used when other features are added to the gateway. The terms 'gateway' or 'portal' have a wider scope and are more general in meaning, while 'hub' stands for a model that offers several services only part of which can be a subject gateway (Koch, 2000a).

model that included a range of services based on an Internet resource catalogue.

There is a substantial difference in the way subject gateways help in resource discovery as opposed to general search services. They are selective with respect to the resources' quality, stability and authority, applying a set of quality measuring criteria and supporting systematic resource discovery. They may be designed with a target audience in mind and they may be created with a certain domain in mind: educational, scholarly, or some particular professional communities or some special field of knowledge or human activity.

From the management point of view, apart from programs for automatic resource harvesting, subject gateways need manual labour to test and improve quality. One of the common features of their management is that they base their resource organization and presentation on resource metadata rather than on the resources themselves.

Some of these services build a comprehensive catalogue of Internet resources based on a widely accepted metadata standard such as DCMES. Earlier services, such as The Online Computer Library Centre (OCLC) NetFirst, contained resource descriptions derived from a bibliographic description standard but extended with elements specific to electronic resources. Apart from descriptive information and summaries, a NetFirst catalogue record contains DDC classification numbers and LCSH. This is because each of these services strives to provide not only reliable subject access to resources but a structured organization and presentation of the information.

The typology of subject gateways may be based on coverage (connected to the subject, geography, language of the resources or type of resource), collaboration with other gateways (shared selection for a common service, record exchange, translation and other co-ordination efforts, cross-browsing, cross-searching, mirroring remote service), and co-operation (co-operation in one subject area, close national co-operation in different subject areas, co-operation of an integrated national service, co-operation between subject gateways and regional or national digital libraries).[24]

---

[24] Kirriemuir (1999), for instance, identified 141 candidates where only 39 could be placed in the category of quality controlled. Koch (2000a) lists 50 subject gateways and quality-controlled subject gateways.

Studies in this area show a tendency to distinguish between two levels of service: **subject gateways** which consist of a list of links with minimal descriptions and shallow subject structure and **quality-controlled subject gateways**, which are subject services based on comprehensive metadata descriptions and resource quality control (Koch, 2000).

Subject gateways have browsing access to Internet resources via a subject structure as their main characteristic. As they may vary in scope, size, granularity, audience, coverage, architecture and technical solutions and source of funding, these characteristics may not be a basis for defining a subject gateway. They are based on resource description but full metadata can be held in a system and only partially displayed (Koch, 2000a).

Quality-controlled subject gateways are characterized by their approach to collection management and development. Although in the initial phase, and later in maintenance procedures, some automatic programs are used, the necessary quality cannot be achieved without human intellectual contribution. Quality-controlled subject gateways perform resource selection based on an established policy and quality evaluation criteria: expected durability of the resource (lifetime), its importance for the user community, the amount of information, size etc. Following these criteria, a collection will attempt to represent a balanced and complete coverage of certain topics, geographical areas or document types. The quality of a service is evident in the rich set of formalised, coherent and standardized metadata for each resource in the collection. A significant part of the metadata is displayed to users. In harmony with standards on a local level, these services will comply with some open standards to allow interoperability with other services. Quality-controlled subject gateways may allow some display adaptability and some user-friendly search support. They are usually developed by libraries, academic and educational institutions or related (there are a few commercial providers, as well, such as OCLC and Ei Village). Most of these services have been developed in the UK, USA, Australia, the Nordic countries, the Netherlands and Germany. Many, however, have global coverage and only a few are oriented towards national resources. The majority of services are in English as one or the only language available.

From the point of view of architecture, these services usually maintain a metadata repository (usually a database management system). Browsable HTML pages for Web display are either static, stored in separate file systems or dynamically produced by program scripts in the search process. When it comes to software support these are very different. The *Resource Organization and Discovery in Subject-based s*ervices (ROADS) project, for instance, provided a package for search gateway metadata management and many services, such as *Social Science Information Gateway* (SOSIG), the *Art, Design, Architecture & Media Information Gateway* (ADAM) and *Engineering E-Library Sweden* (EELS) use this package and store resource records in the database. An alternative to this is the OCLC software MANTIS and SiteSearch.

Interoperability, in the sense that it has to support cross-gateway searching and browsing, has to agree upon the level of protocol, query language, record format and syntax, metadata schema, subject classification used and cataloguing rules applied. Cross-browsing between different classification schemes is not easy to achieve and involves much time-consuming labour on classification scheme mapping. Techniques and methodologies used in this area are scarce. One example of mapping between DDC (Biz/ed) and UDC (SOSIG) is, for instance, explored within the ROADS information gateway (Hiom, 1998).

The number of *subject gate*ways and *quality-controlled subject gateway*s is growing, and existing entities are constantly changing in scope, coverage, content, methods and features (Koch, 2000a). The overall trend is towards brokerage services based on distributed communication components. Dempsey (2000) points out that the future of these services will depend on their ability to move to some sustainable funding source and that it is not evident whether they are going to retain their institutional identity or whether they are going to be absorbed into a national learning service, professional portal service, etc.

There is a tendency to bring these gateways together in a new, federated structure, which will bridge interdisciplinary and cross-sector information gateways. *Resource Discovery Network* (RDN) is a follow-up of the *Electronic Libraries Programme* (eLIB) subject gateways, and is a network organization with a centre and a number of organizations (hubs) e.g. SOSIG (social sciences hub), EMC (engineering, maths, computing hub) and Humbul (humanities hub).

The field of quality-controlled subject gateways is of particular importance for library classification. There are, in effect, numerous developments in the area of quality information services which are connected to a subject approach to information and the deployment of library classification to achieve better cross-subject and cross-language resource discovery. Many efforts are directed towards standardization of the procedures and cross gateway searches and a number of projects are being funded to merge the efforts scattered across individual services. The use of classification, and UDC in particular, in quality-controlled search gateways will be discussed in more detail in Chapter Three.

1.4.3 Semantic Web and KOS

On the subject of resource discovery, in 1996, Berners-Lee pointed out that information is on the Internet but is hard to discover unless put in a form that is actually a semantic statement i.e. some knowledge representation language statement (Berners-Lee, 1996). In subsequent years he announced the idea of a global reasoning Web at the WWW 7 (1997) and the WWW 8 (1998) conferences when he formally introduced his vision of the Semantic Web. Information discovery enters a new dimension, as the objective of the Semantic Web is to link substantial parts of human knowledge and allow them to be processed by machines.

The semantics of subject metadata and their representation in a contextualised and machine-understandable terminology are the key elements. The idea of a Semantic Web, as expressed by Berners-Lee, is that machine-processable, ‘meaningful’ metadata will be the basis for a new generation of information retrieval services that will help both humans and computers to access information and communicate with one another. This will enable, for instance, intelligent agent programs to operate and fulfil more demanding tasks independently (Berners-Lee, Hendler & Lassila, 2001).

While, at present, there is information on the Web for the human reader that can be navigated by a simple link, in the future, data will be processed by programs designed independently of data, i.e. using 'domain independent' reasoning processes. These programs will require machine-readable statements about resources and their relationships depending on:

i) the existence of a common model
ii) a link between vocabulary terms and their unique definitions
iii) the availability of definitions that can be accessed by programs

The vision implies that software agents will be navigating a Web that consists of descriptions and “ontologies” (including unknown vocabularies), making inferences about the data collected and communicating via partial understanding. The implication of the Semantic Web is that the Internet would be searchable not only using words but also through meaning. As Veltman pointed out (2001, 2002) this obviously requires both semantic and syntactic interoperability of a subject vocabulary, since subject description is based on both conceptual isolates and propositional logic.

In this context, existing KOS such as classification systems have been recognized as important sources of structured and formalised vocabularies that can be used to support the development of the Semantic Web. In this context Soergel (1999: 1119) summarises classification's rôle as:

- providing semantic road maps; improving communication and learning
- providing a conceptual base for the design of research
- providing classification for actions
- supporting information retrieval
- providing a conceptual base for knowledge-based systems
- providing the conceptual basis for data element definition and object hierarchies in software systems
- cross-discipline, cross-language and cross-culture mapping
- serving as a mono-bi-or multilingual dictionary/knowledge base for natural language processing

Regardless of the indexing term used (i.e. notational symbol or word), classification is recognisable by the logical processes involved, its structure, or by the knowledge representation functions performed. Information system implementers look for classification structures that can be used in information mapping, concept mapping, visualisation of subject access, interactive search presentation and distributed resource viewers. Each one of these applications is closely related to the availability of classification data in a machine-processable form.

Two developments in the area of KOS, and in particular of classificatory knowledge structures, are seen as a way of supporting the idea of the Semantic Web and they are likely to influence the future use of library classification:

i) standards and applications for terminological exchange
ii) ontologies (as developed in the field of artificial intelligence)

Standardization is primarily focused on a technological Web framework and a move towards Web-based representational languages (*Extensible Markup Language* - XML and XML/RDF- *Resource Description Framework*).

1.4.3.1 Ontology

One focus of the Semantic Web is the development of machine-processable knowledge structures, which connect the Semantic Web and the fields of knowledge engineering and AI. Information system implementers are working on building or adapting a machine-understandable and shareable vocabulary using standardized and shareable formats as developed by the *World Wide Web Consortium* (W3C) or other fora (Noy & McGuinness, 2001).

The importance of *knowledge-based systems* (KBS) has been analysed in information science since the 1980s, mostly in relation to automatic indexing (natural language processing) and information retrieval (Croft, 1989). As text processing has become central to information discovery on the Internet, specific concepts such as *knowledge domain* and *ontology* are used more frequently in information science (Vickery, 1997).

The term 'ontology' itself begins to embrace an entire set of meanings and comprises everything from taxonomical categories, controlled vocabularies in resource metadata, lists of products or classifications of services, database vocabulary and relationships. An ontology, in the sense of a *formal data structur*e used to build a *knowledge base*, was introduced in the 1990s and is a relatively new research topic (Ding, 2001; Vickery, 1997a).

It is suggested that an *ontology* works best if understood to be a formal explicit description of concepts in a domain of discourse and to be a source of shareable semantic data. An ontology is closely related to KBS. These expert systems have to be 'fed' with knowledge from a particular domain (*knowledge base*) and programmed to perform procedures (part of a program called an *inference engine*) to solve the task. In order to achieve this, a *knowledge base* has to be built on principles of formal machine-processable data structures.

A *Knowledge base* itself is actually an informal term for a collection of information that includes an *ontology* as one of its components. It also may contain, however, information specified in a 'declarative language' such as logic or 'expert system rules', or non-formalised information expressed in a natural language or procedural code. An ontology is built following a basic logical procedure and this results in a classification structure with clearly defined classes and conceptual relationships that, for instance, can be expressed through formalised structures called 'conceptual graphs' and formatted in a machine-processable way (Sowa, 2000).

Knowledge representation as understood within the field of AI deals with a wide range of knowledge that is computable i.e. expressed by strict rules of logic. The expressive power of logic, as pointed by Sowa, includes every kind of information that can be stored or programmed on a computer. Logic, however, has no power to describe things that exist. Logic itself is a simple language with a small number of basic symbols, thus the predicates that represent knowledge about the real world have to come through ontology.

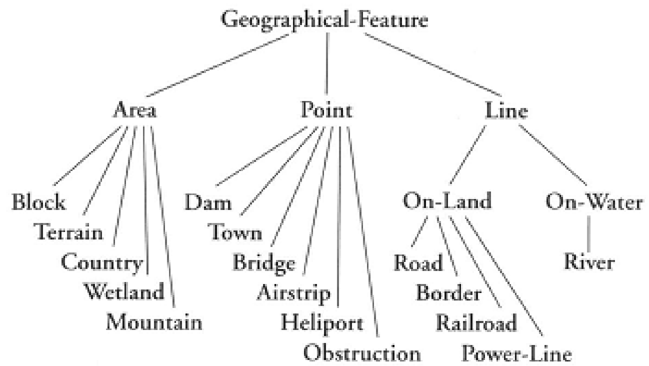


Figure 1.2: Graphical representation of a 'micro world' ontology

Ontological categories are collected through observation and reasoning that provide information about abstract and concrete entities and their types and relationships in a particular domain. It could be said that an ontology is a study of the categories of things that exist or may exist in some domain. The product of such a study is a catalogue of the types of things that exist or are assumed to exist in a domain of interest from the perspective of a person who uses an agreed language for the purpose of talking about this domain, as illustrated in Figure 1.2 (from Sowa, 2000: 53). The logic combined with an ontology provides a language that can express relationships between entities in a domain of interest.

In the context of the Semantic Web, an ontology can have a very broad meaning usually based on the classification structure and vocabulary control that are inherent in every ontology.[25] Helpful categorization in terms of the practical application of ontologies that reveals this link to classification is given by McGuinness (2002), who draws a distinction between *simple ontologies* and *structured ontologies*.

S*imple ontologies* are taxonomies or simple hierarchical vocabularies.[26] According to the way she defines their purpose, it is obvious that these ontologies are used in the same way as any library classification would be used:

- to provide a controlled vocabulary
- for site organization and navigation support
- as an "umbrella structure" from which to extend content
- for browsing support
- for search support
- for sense disambiguation support

Structured ontologies, on the other hand, apart from being machine-readable encoded hierarchical relationships, contain information about properties and value restrictions of the properties. For instance, a class 'goods' can have the property 'price' whose value could be restricted to numbers or a number range. As a concept in an ontology is described in term of classes, properties and rôles, and these are encoded to be machine-readable, any part of the concept encoded structure can be more specifically defined in terms of values that can be associated with it. Because of this quality, *structured ontologies* could be used as a part of an application environment to help with:

- consistency checking
- completion (automatic inclusion/exclusion of properties)

---

[25] Examples of ontologies from the fields of linguistics and knowledge engineering are: WordNet (structured vocabulary of the English language), Cyc (knowledge base and inference engine with a general common ontology), SENSUS (a linguistic domain ontology) and STEP (Standard for exchange of product model data - ontology built to exchange product data among different computer systems).

[26] Examples: DMOZ http://www.dmoz.com and UMLS - unified medical language system at http://www.nlm.nih.gov/research/umls.

- interoperability support (obtaining missing information through the link to other properties)
- encoding test suites
- configuration support
- support, structured and customized searching
- exploiting generalization/specialization

Ontology representation languages are standards for the machine-processable encoding of an ontology. They supply an expert system with a complex ontological framework, domain concepts and reasoning rules. Recent research activities have focused on establishing necessary standardization in this area and today's ontology language encoding standards are trying to merge language expressive power with reasoning power to provide a representational language with known reasoning properties (McGuinness, 2002). Some alternative approaches for modelling ontologies have emerged from the field of software engineering. These are based on modelling constructs, analysis and design of object-oriented software systems (Cranefield, 2001).

Languages used to represent ontologies belong to three categories: logic-based (first-order logic), frame-based (frame logic) or Web-based (RDF, XML, HTML). While the first two are particular to AI applications, Web-based ontology representation languages could be used, to some extent, to support the representation of vocabularies outside the AI domain. Three Web ontology languages: *DARPA Agent Markup Language (*DAML), *Ontology Web Language* (OWL) and *Simple HTML Ontology extension (*SHOE) are briefly described in Appendix 1.1.

1.4.3.2 Interoperability among KOS

Developments in the area of Web ontology languages have important implications for the understanding of the future use of traditional KOS such as classification. In order to support 'one stop' seamless resource discovery across distributed collections on the Internet, knowledge organization tools need to be used in a more automated manner. So as to support alternative subject access, KOS tools ought to support different kinds of mapping, switching, translation, adaptation and linking through temporary union lists. This means that KOS tools, when used in the networked environment, have to be represented and made available in a way that they can interact and be exchanged between machines (Zeng & Chan, 2004).

Important steps have been already made to provide common standards for representation of indexing languages in order to build *terminological services.* The purpose of terminology servers is to provide access to mappings between different controlled vocabularies and thus act as a searching mediation service between distributed systems (Vizine-Goetz, 2003; Vizine-Goetz et al., 2004, 2004a; Davies Ron, 2004). Prerequisites for such a development are open and platform independent vocabulary encoding standards. Soergel (2001: 1) defines the purpose of these standards as:

i) input of KOS data into a program and the transfer of data from one program to another
ii) accessing KOS for applications, querying KOS and viewing results
iii) identifying specific terms/concepts in specific KOS
iv) prescribing and giving guidance on good practices

New developments in the wider Internet community include efforts to make use of existing data by providing a platform for the linking and exchange of different indexing languages looking for common representations and protocols (Tudhope & Koch, 2004; Zeng & Chan, 2004; Chan & Zeng, 2004). One of the active initiatives in this area is the *Networked Knowledge Organization Systems/Services* (NKOS) which acts through workshops, publications and a mailing list and brings together implementers and standards developers from different domains.

Standards already developed in specific information sectors and domains (librarianship, digital libraries, geographical data, government data, archives, e-learning) are now being compared, evaluated and tested using more transparent and flexible data transport standards such as XML and XML/RDF. Such is the case for standards for the machine-processing (translation) and exchange of dictionaries, glossaries and concept maps as well as those already created for the exchange of thesauri and classifications. Existing standards for transport and exchange of KOS fall into the following groups:

i)Standards for terminology exchange (dictionaries, glossaries):

- Open Lexicon Interchange Format (OLIF)
- ISO 12200:1999 - MAchine-Readable Terminology Interchange Format (MARTIF)

- ISO 16642:2003 - Terminological Markup Framework (TMF)

ii)Standards for library KOS

- MARC Authority Formats
- MARC 21 Concise Format for Classification Data
- UNIMARC Format for Classification Data
- The Zthes profile

iii)E-learning vocabulary standards

- Vocabulary definition exchange (VDEX)

iv)General (not domain specific) standards for vocabulary exchange

- XTM - XML specification for ISO/IEC 13250:2000 Topic Maps
- The Vocabulary Markup Language (Voc-ML)
- eXchangeable Faceted Metadata Language - XFML Core
- SKOS - Simple Knowledge Organisation System (in progress)
- BS 8723 - Structured vocabularies for information retrieval - guide (in progress)

More information on each standard is given in Appendix 1. When directed at the use on the Internet, the majority of these standards will require XML or XML/RDF encoding. The application of any of these standards is not always straightforward. This is especially the case for an existing controlled vocabulary, in which case the encoding schema used needs to be adapted so as to prevent data or functionality loss. From the above-mentioned groups, *Standards for library KOS* include formats especially created for library classification, however, some issues remain unsolved and these will be explained in more detail in Chapter Six.

1.5 Literature review summary

UDC's analytico-synthetic structure allows greater indexing specificity. This was the reason for its wide application in special libraries and bibliographic databases, where it has been used for online IR since the 1960s. The use of classification in IR up to 1990s was often based on searching UDC numbers. Important requirements for the successful application of searching UDC online that evolved during this period were, however, the following:

i) the need for post-coordinate searching of pre-synthesised elements

ii) the importance of not relying on UDC symbols in searching i.e. it was necessary to code UDC numbers in order to search for elements of pre-synthesised numbers (this was demonstrated in both the Freeman & Atherton's and in Caless' projects)

iii) the importance of making provisions for searching UDC using words

Following a period of neglect in the development of the classification in the 1980s, the UDC ownership and management changed in the 1992. The schedules were automated in 1993, which made their management and revision easier and their publication and implementation faster. This was criticial at a time when interest in classifications within the library community was at its lowest.

IR trends indicate greater flexibility with respect to the application of classifications, their combination with other indexing languages, as well as their use in conjunction with advanced retrieval techniques. Modern interactive IR systems endorse the view that a combination of searching and browsing and the transition from one to the other are fundamental in supporting users' information behaviour. In order to support subject browsing some predefined classificatory knowledge structure is paramount. For use in a modern information retrieval environment, classification has to be implemented with flexible, multidirectional browsing function and be accessible through natural language. These functions depend on the way in which an IR system can process classification data.

Resource discovery on the Internet places an emphasis on subject retrieval and has contributed to a revival of interest in library classifications. General classification systems are used on the Internet for subject browsing and cross-collection/cross-domain resource discovery. In conjunction with automatic indexing, using classification saves human effort by completing tasks that are impossible to perform manually on large resource collections but it also allows for subject collocation of previously classified legacy bibliographic data with a constant influx of resources that are created digitally.

The complexity of the use of classification depends on resource metadata which is the carrier of classification data. A machine-processable metadata infrastructure is essential for the development of the Semantic Web, which is the reason why vocabularies used in metadata description are now being formatted and encoded in a standard manner in order to make them easier to process and be exchanged by machines. The use of classification on the Internet is closely dependent on a selected vocabulary's ability to be expressed in a more standardized and machine-understandable way.

CHAPTER TWO: RESEARCH QUESTIONS AND METHODOLOGY

2.1 Research problem and research questions

Background research into the trends in IR and information discovery on the Internet indicates that classificatory indexing languages play an important rôle in resource discovery and that this is likely to be even more true with Semantic Web developments. Metadata standards and standards for KOS transport and exchange are being created as an infrastructure for information organization and discovery on the Internet and these standards envisage and facilitate the use of library classifications.

The main problem explored by this research project is the poor use of classification systems, UDC in particular, in information retrieval. One of the reasons is the inadequate machine-readability of classification data, which also affects the management and exchange of a classification in the networked environment.

The use of classification in a networked environment depends on both standardization of the management of classification data for use in IR and standardization for its exchange and sharing. The goal is to satisfy these two functions with a single standard classification format.

The problem with the use of UDC in IR is twofold. Firstly, poor exploitation of the UDC in IR stems from the failure of systems to support the hierarchical and synthetic features of the UDC. Secondly, UDC data are distributed from the database, which is created for publishing and management purposes and for holding data with limited machine readability. In order to support IR functions data have to be enhanced and additionally encoded.

The specific focus of this research is to make a connection between the way the UDC is used for information retrieval and the way it is encoded for machine processing. The recommendation would be to have properly structured machine-readable classification data at source. Such data would support all the applications and functions that are built on top of them and would make classification implementation and exchange easy and cost efficient.

2.1.1 Research scope and research questions

This research especially focuses on the factors that would determine and facilitate the future use of UDC in the networked environment. Present research builds upon important suggestions in relation to the improvement of machine-readability of UDC data and its use in IR which have been outlined since 1990 (Buxton, 1990, 1993; Loth, 1996; Riesthuis, 1998b, 2003). Key assumptions behind the research are that:

- the exchange of information in a networked environment relies on a high level of computer readability. The way to achieve this is through a formalisation and standardization of information structures
- the use of any KOS for resource discovery on the Internet depends on its availability in a machine-readable format. This format, however, has to be both standardized (i.e. system and platform independent) and capable of supporting classification-specific functions. The prerequisite for the creation or use of such a format is the proper understanding of the UDC structure, data representation, data elements, their relationships and how they are used in IR
- current trends in IR and the current development of Web technology determine the direction for the development of classifications and will ultimately decide their future

This research examines the way the UDC's structure and vocabulary underpin indexing and information retrieval. In relation to this, attention is paid to the kind of classification data formatting and management necessary for this to be exploited in a networked environment.

Although the use of classification in a networked environment is also dependent on the economy of its ownership, management, distribution and market demand, this research mainly focuses on the improvements that might be achieved through better system availability for machine processing. It is beyond the scope of this research, however, to suggest any particular standard for classification modelling or encoding, as such a proposal is only relevant if made by standardization bodies or demanded by classification owners.

This research will attempt to answer the following specific questions:

- What evidence is there of UDC quality in information retrieval and what is the evidence for its suitability for knowledge organization and information exchange in the global network?

- What are the automation issues and the requirements (in a networked environment) regarding the full exploitation of the UDC vocabulary, vocabulary management, data format and classification data quality?

- What are the areas in need of improvement and what are the specific recommendations for facilitating the use of the UDC system in a networked environment?

2.1.2 Possible benefits

For the management and maintenance of the UDC. Summary evidence of existing UDC use, functionality and application in the networked environment will highlight users' needs with respect to the form and format in which classification data are distributed and shared. The requirements with respect to UDC data may lead to enrichment of classification data, which may influence its future development, distribution and implementation. Standards and formats for the encoding of classification data can be adapted and used for the UDC, provided there is clear understanding of the functions UDC may support in IR and a proper data model that respects UDC's own features. The formalisation of UDC data that is necessary in the new networked environment may also have positive effects on the system itself by making it more consistent and coherent.

For the users. The research results may help to improve the UDC data structure that will facilitate UDC implementation and its use in subject browsing and searching. The transparency of the vocabulary structure and its functionality may encourage the export of the UDC to some standard Web vocabulary exchange format and its mapping to other systems. UDC is one of the most widely used KOS in the world and a greater presence on the Internet may ease cross-collection subject access to legacy data worldwide. Access to and exchange of UDC data via an independent and standardized Web encoding language is likely to encourage its use among new stakeholders looking for a general KOS to support browsing.

For the knowledge organization research and standardization. The UDC is, at present, the only general analytico-synthetic classification that is automated and is widely used for information retrieval. It contains data that exemplify purely faceted structures, its vocabulary organization is relational and its synthetic features correspond to those exhibited by fully faceted classification systems.

UDC syntax allows the expression of different relationships between subjects and has the qualities of a powerful pre-coordinated indexing language. The UDC possesses all the elements that typify an advanced classification system and its data model may be used as a starting point in researching any other classification of this type or creating standards and formats. This may help in advancing towards a generic ontology for processing a faceted and pre-coordinated vocabulary with a classificatory structure, which is still missing in the KO community.

2.2 Research methodology

This research deals with issues relevant for classification automation during the period 1990 to 2005. It focuses on the interdependence between UDC functionality in retrieval and the format of UDC data available for machine processing. The main objective is to investigate the circumstances of the use of classification in a networked environment and analyse them with respect to the known and standard use of classification in IR and the available forms of the classification data structure. The relationships between the existing format of UDC data and format for management and its use in a networked environment are addressed more specifically. The research developed in five steps which are represented as five related but independent studies.

2.2.1 Procedure

The first step in this research was to investigate the presence of classifications, the UDC in particular, and their rôle in knowledge organization on the Internet (Chapter Three). Subject gateways and research projects implementing classification were observed and an account of the functionality of the search services using UDC was given (Section 3.4). The nature of these services and trends in the use of the classification were established. This is followed by a study of metadata standards recommending the use of classification in subject description along with trends and developments in this area (Section 3.5).

The investigation of trends and the development of KOS in a networked environment was based on the literature, observation of search services and standard bodies and fora involved in metadata creation, but was also based on participation in workshops and discussion lists. Practical experience gained through participation in two research projects that involve UDC implementation in metadata (Educator Access to Services in the Electronic Landscape project - EASEL - 2000-2002) and the creation of a faceted classification for a humanities portal in the FATKS (2002-2003) project was especially relevant in providing background for this research.[27]

The second step was a study focused on specific issues of the UDC, its performance and its functions, in IR (Chapters Four and Five). This included a summary of known and well researched uses of classification in IR for the purpose of query disambiguation and contextualisation and for improving precision and recall. Functions such as searching and browsing UDC were examined. Special attention was paid to the difference between searching the classification using notation and searching using words. A summary of the interface issues in classification online applications was addressed in Chapter Five. An overview of functionalities of Web OPACs was provided, through the analysis of 30 different catalogues, demonstrating different levels of use of the UDC in collection searching and browsing.

The third step was the study focusing on UDC practical use and the way UDC data were formatted for machine processing (Chapter Six). Special issues in automation related to the encoding of synthesised numbers and hierarchical and associative relationships were mentioned. The classification authority file as a tool for the management and exchange of classification data was proposed and two standard MARC formats for the creation of these files were studied. The analysis of the UDC in IR was based on a study of the literature on UDC applications. This was synthesised with studies and research on the use of other classification systems and KOS in IR systems during the 1980s and 1990s, wherever these contributed to the general understanding of the use of classification in IR or the management of classification data.

The fourth step (Chapter Seven) was the study reports on a UDC user survey and UDC translation into world languages. Following this, an overview of UDC

---

[27] Project web site at http://www.fdgroup.com/easel is no longer accessible. Project facts can be found at http://dbs.cordis.lu/fep-cgi/srchidadb?ACTION=D&SESSION=26452005-1-16&DOC=5&TBL=EN_PROJ&RCN=EP_RPG:IST-1999-10051&CALLER=PROJ_IST. Project Facet Analytical Theory Knowledge Structure (FATKS), 2002-2003 can be accessed at http://www.ucl.ac.uk/fatks/index.htm.

management and maintenance was made describing the present UDC database and UDC MRF update and distribution. This report reveals the extent to which this database format meets the requirements of managing and exchanging UDC data in its current applications. An analysis of the UDC management model was given, focusing on its relationship to the current use of classification in IR systems. Information on the revision, update, management and distribution was collected through participation in the Revision Committee of the UDC (1999-2004).

The last step in this research looks into the relationship between UDC MRF data quality and the database format (Chapter Eight). The study starts with the validation of the UDC MRF 2003, highlighting issues of UDC data access and notation validation. This validation indicates the requirements for enhanced data formatting. The proposal for a new database format, put forward by Riesthuis in 2003, was examined in relation to these requirements. Following this, a UDC conceptual data model was proposed which would be more compatible with emerging standards for KOS application and exchange in a networked environment. An entity-relationship model and data element schema developed for the *UDC Research Database* were provided which exemplified a data format that may better suit requirements for maintenance and use.

The conclusion of the research (Chapter Nine) summarises the sequence of arguments for improved and more standardized classification data modelling. It highlights the dependency between the use of classification and the availability of data for machine processing and gives recommendations for improvements in the management of classification data.

2.2.2 Materials and techniques

This is qualitative research that combines seven smaller primary studies with a literature study. The primary studies were relevant in collecting data for analysis of UDC use, users, its current application on the Internet, UDC management and the quality of its data. Methodological (literature studies) were relevant for establishing a wider context of classification use and trends in its application.

2.2.2.1 Primary studies

There were three types of primary study involved in this research: observations, survey and empirical work.

Observations. The use of UDC in Internet subject gateways was observed. Nine gateways with an English interface that appeared between 1990-2004 were described and information was given on date, coverage, type of indexing and language interface (see Chapter Three, section 3.4, Table 3.3). Seven gateways using UDC were then examined in more detail and eight interface functions (hierarchy display, hierarchy level, caption, number synthesis, number search, search to browse, use of UDC MRF) were compared. Summary results are presented in Chapter Three, section 3.4.3, Table 3.4. Screen captures of these gateways are given in Appendix 2.

Five metadata standards (DCMES - Dublin Core Metadata Element Set, LOM - Learning Object Metadata, AMICO - Art Museum Image Consortium, EAD - Encoded Archival Description, MARC - MAchine Readable Cataloging) were described from the point of view of their ability to accommodate and transport UDC data (Chapter Three, Section 3.5, Table 3.6). Examples of metadata instances are given in Appendix 3. The summary of metadata schema characteristics looked at were: the number of schema elements carrying a subject description, occurrence (repeatable/non-repeatable), identifier of a classification scheme, identifier of a concept, linking to external vocabulary, simple or structured syntax for a UDC notation, and reported use of UDC. These were presented in Chapter Three, section 3.5.5, Table 3.7. Information on the use of classification in metadata was collected through literature and standards study, observation of metadata standards development, participation in metadata workshops and work on the EASEL project that involved the use of UDC in LOM metadata.

Thirty Web OPACS using UDC from twenty-two countries were examined. Countries were selected based on their usage of the UDC (see Chapter Five, section 5.3.2, Table 5.4), while libraries and OPACs were selected randomly from known and available users. The list of OPACs and their details is given in Appendix 4.1, Table 4.1. Through initial observation, a selection of twenty-three UDC based IR functions (see Chapter Five, Section 5.3.2.1) was identified and checked against the selected OPACs. Functions were analysed and compared across different catalogues and summary results were given from the point of view of what kind of system a library used (vendor/in-house) and for what purpose UDC was used (shelf arrangement or indexing).

Surveys. A worldwide UDC user survey was conducted in 2004 and included librarians in 208 countries (Chapter Seven, section 7.1.1). The materials are presented in Appendix 8: the interview design diagram in section 8.1, a sample of the initial e-mail in French, Spanish and Portuguese in section 8.2 and a sample of responses in section 8.3.

The UDC user study combined techniques of an e-mail interview with literature research and was designed to establish the number of countries in which UDC is used and to what extent. The objective was to collect up to date information about the use of UDC in countries for which there was no written evidence of use and to verify the type of usage in countries known to be using UDC. Because of the lack of recent studies on UDC use worldwide and differences in UDC use in different countries it was not possible to approach all professionals with the same kind of questionnaire.

The technique of e-mail interview was considered especially suitable because of its interactive value, the possibility of clarifying ambiguities, for reaching subjects that are otherwise hard to locate, but also because this technique encourages more detailed and carefully considered answers (Bampton & Cowton, 2002; Hewson et al., 2003). For this particular study the technique was chosen because it allowed respondents to be more accurate in answering factual questions and to go back and check information, which enhanced the validity and quality of data obtained.

Persons for contact were randomly chosen from different categories of librarians (cataloguers, management staff in public, special and academic libraries and LIS teaching staff). Coverage was the most important goal and the e-mail interview allowed the research to explore the most appropriate approach, acquire information on further contacts and change the language of correspondence or questions according to the situation. Information initially obtained was used as the basis for correspondence with other professionals. In order to obtain a response, an average of five e-mails (not counting failed deliveries) with an initial question were sent to 208 countries. On average an interview consisted of the exchange of up to 6 brief e-mails. Data collected through the interview were then compared with literature research.

The UDC bibliography and translation survey also combined literature research with an e-mail interview technique. Survey results were compared with literature sources and summary results and analysis is given in Chapter Seven, section 7.1.2 where the method of data analysis was also given. Results of the UDC translations bibliographic survey involved the research of existing bibliographies, e-mail enquiries and verification in the library catalogues of relevant countries. After the literature survey was conducted, the last reported translations were summarised in Table 7.4. The full bibliography of the last reported translations was listed in Appendix 9, section 9.3.

Practical experience and experimental work. This includes UDC MRF maintenance in 2000 - i.e. practical work on the MRF update which was relevant for the observation of processes and procedures involved in UDC database maintenance and update. The 2000 annual update of the MRF was completed based on the instructions given in the Master Reference File manual.

In addition, practical work on UDC revision (as a member of the Revision Committee team from 1999-2004) was relevant for observing the interaction between revision policy and procedures and how these affect UDC database management, and also how the existing database supports the revision work. Observations on this are summarised in Chapter Seven.

The validation of UDC data was based on an investigation of the UDC MRF 2003 database containing 66,733 records. This was summarised in Chapter Eight. A separate relational MySQL database was created by the author of this thesis and a software developer, called the *UDC Research Database*, in order to validate fully the UDC MRF 2003 data. A data element schema of the database is given in Appendix 10, section 10.3.

2.2.2.2 Secondary (literature) studies

Selection of the literature and a methodological study followed different patterns in relation to the following four subject areas:

- UDC
- IR - general issues
- indexing and IR (interface, OPACs, authority control)
- KO on the Internet (subject gateways, metadata, Semantic Web)

The selection of literature analysed in relation to the UDC was exhaustive and covered all traceable literature on UDC automation and retrieval issues and also UDC structure and data representation. Care was taken to include research reports from the period 1990-2004 where significant changes in the technological environment took place. Literature on this topic included monographs and textbooks on the UDC, journal articles and conference proceedings, project reports (published and unpublished) and internal UDC Consortium materials. General issues in IR trends and developments were studied and the choice of literature included only a small number of relevant, frequently cited textbooks and articles.

With respect to the relationships between indexing languages, classification and IR, the selection was biased towards research reports and surveys on the use of classification in IR. The initial set of literature was selected from the proceedings of the eight (international) conferences of the *International Society of Knowledge Organization* and six *International Study Conferences on Classification Research*. In addition, a quantity of relevant literature in LIS journals was discovered through bibliographic citations and tracing related research or following the work of researchers involved with classification and IR such as Pauline Cochrane, Karen Markey Drabenstott, Elaine Svenonius, Charles Hildreth and Winfried Gödert.

Literature on knowledge organization on the Internet consisted predominantly of journal articles and conference proceedings, project and workshop reports and standard fora documentation. The selection of these was based on the quality, authority and relevance of the topic. The observation of standards bodies and participation in working groups on metadata and discussion lists was another source of useful literature that supported this study.

The literature used is predominantly in English with a small number of books and articles in French, German, Italian and Croatian. Language bias is an especially sensitive issue for the part of the research that involves UDC as its use in IR is much more advanced in non-English speaking countries. An attempt was made to alleviate language bias by extensive coverage.

CHAPTER THREE: CLASSIFICATION ON THE INTERNET

3.1 Introduction

Interest in library classifications increased in parallel with the recognition of the importance of subject access in resource organization and discovery on the Internet. In the 1990s classifications have become the KO tool of choice for many academic and research orientated resource discovery services. This chapter starts by listing the major research projects responsible for the development of information services based on library classification, followed by an explanation of the context in which library classifications were selected and evaluated. An overview of *quality subject gateways* using UDC with an English interface, in the period 1992-2004 is given. The observations concentrate on the interface rather than on the underlying IT infrastructure with special attention paid to quality subject gateways with automatic indexing.

The chapter closes by exploring metadata standards as a fundamental infrastructure for the use of library classification in organization and resource discovery. Various widely used metadata standards are analysed in relation to their rôle as subject data carriers and their potential to accommodate a classification was observed. Five metadata schemas, one general and four domain specific, are compared: DCMES, LOM, AMICO, EAD and MARC. Issues relating to the use and usability of classification in metadata with regard to Web architecture are addressed.

3.2 Research in classification

Free text searching and purely linguistic methods proved inadequate for information management and discovery in large collections. Library classifications are multilingual, hierarchical indexing systems and have been recognized as valuable in improving resource discovery on the Internet. Their potential has been examined in numerous papers (Vizine-Goetz, 1996, 1997, 1998a; Koch et al., 1997, 2000; Ardö et al., 1994, 2000; Chan, 2000).

A number of research projects explored the rôle of classification and its adaptation to resource discovery in the networked environment and these had an important impact on the use of classification and its popularity outside the library domain. Research projects were also important for the development of library classifications, as they provided an architecture, infrastructure and model for classification use and sharing. Their specific focus was on the following:

- development of an end-user interface for displaying classification; hierarchy structures
- transformation of classification captions to an end-user language
- linking classification to other subject languages
- building the classification authority files (or thesaurus, or vocabulary)
- automatic document indexing programs using classification

DDC was the most studied and used classification because of an exceptional situation at the OCLC that both holds the copyright to the classification and conducts research to explore the use of the classification in resource discovery. Although these experiences are associated with one classification scheme, they are significant for other classification users and developers. Based on experience with DDC and OCLC research, Mitchell (1996) pointed out that new uses of classification are driving structural and content changes in the underlying DDC database, and classification schemes ought to change to accommodate new information needs. Vizine-Goetz reported on the limitation of the DDC vocabulary and concluded that, not only is the extension of traditional classification schedules needed, but also links between classification notations and search terms beyond the scope of a traditional subject alphabetical index are required. Also, the need for a hypertext interface has been clearly expressed (Hickey & Vizine-Goetz, 1999; Vizine-Goetz, 1999; Vizine-Goetz, 1998).

3.2.1.1 OCLC research projects

OCLC projects exploring the potential of library classification were initiated as early as 1991 and are still ongoing: 1991-1993 Internet Resource Project; 1993-1996 InterCat; 1996 NetFirst; 1997 Scorpion; 1999 *Cooperative Online Resource Catalog* (CORC) and 2003 projects on terminological services that explore the potential of open software, such as the *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) and Web Services, in enhancing the exchange and sharing of subject vocabularies.

The first two projects, the *Internet Resource Project* and *InterCat* introduced and explored the practice of cataloguing Internet resources. The aim of the *NetFirst* project in 1996 was to build an authoritative directory of Internet-accessible resources (available through the OCLC *FirstSearch* service). The records contain detailed bibliographic information, summary descriptions, subject headings and DDC classification codes. The database coverage is: Web pages, Gopher servers, electronic discussion lists, library catalogues and electronic publications. *NetFirst* was reported to contain 120,000 records of Internet resources (Cross, 2000a). The service is still active and records are accessible through the OCLC *CORC* service.

The *Scorpion* project initiated in 1997 aimed at building tools for automatic subject assignment, combining automatic indexing techniques and DDC (http://orc.rsch.oclc.org:6109/). *Scorpion* was envisaged as an aid to human cataloguing by automating subject assignment where items are available electronically (Shafer, 1999). Each electronic document to be classified was treated as a query against the DDC database using ranked retrieval. The result of the search can then be treated as the subject of the document, although cataloguers were reminded not to rely completely on automatically assigned classification numbers (Thompson, Shafer & Vizine-Goetz, 1997).

In 1999 OCLC extended their co-operative model used to build *WorldCat*, to Web resources via the *CORC* project. *CORC* provides tools for the co-operative creation, maintenance and use of metadata for Web resources and uses previous project outcomes such as the tool for automatic classification developed in *Scorpion* (CORC, 1999). The project attempts to provide access control for names or classification numbers in cases where different content can be supplied for different types of user (e.g. the meaning of a classification number could be given in different languages, or related to different user ages or education levels). *CORC* also explores an XML/RDF architecture in order to enable access and make use of existing authority files rather than creating new authority records.

CORC records link DDC numbers and LCSH and this linking gathers a number of related subject headings under one classification number. *CORC* is built on the co-operative creation and sharing of metadata by libraries and the co-operative cataloguing of Web resources. It accommodates both local and shared metadata, promotes the mixing of metadata for physical and digital items, authority control, RDF/XML import/export, portal page import/export, integration of Dublin Core and

MARC into one system, flexible harvesting of resources, assisted classification and subject heading assignment, automatic keyword extraction, link maintenance and access to Z39.50 browsing interfaces (Hickey &Vizine-Goetz, 1999).

Since 2003, OCLC developments have been moving toward terminological services that provide shareable vocabulary data using open vocabulary encoding standards, XML and Web services. Within these initiatives there are a number of significant developments in vocabulary mapping: *(a)* linking of DDC and LCSH, *(b)* linking LCC, National Library of Medicine classification (NLM), DDC *(c)* linking Genre terms for fiction and drama (GSAFD) to parts of LCSH and *(d)* ERIC thesaurus to LCSH (Vizine-Goetz et al., 2004).

3.2.1.2 European research projects

ROADS - Resource Organisation and Discovery in Subject-based Services (http://www.roads.lut.ac.uk/) was an UK project. The main objective was providing a set of software tools for the UK eLIB subject-based services (http://www.ukoln.ac.uk/services/elib/). It provided a collaborative framework for the UK community through which it would contribute to the design and development of quality information services. The initial intention was to have unified access to distributed resources to spare the user from the fragmentation found in current bibliographic practice.

The project promoted the concept of a "trusted information provider" based on the assumption that the collection descriptions from approved authors and Web-site administrators were essential to a sustainable service. This was facilitated by a simple record structure that enabled easy data creation and gathering (Dempsey, 1996). The project was focused on providing an Internet standards-based application for the creation of resource descriptions, which was simple but informative enough to support effective retrieval. For this purpose *ROADS* prepared a number of templates for the management of an information gateway.

DESIRE - Development of a European Service for Information on Research and Education (http://www.desire.org) was a large pan-European project started in 1996 with the aim of promoting the use of the World Wide Web in the European Community. Two project phases *DESIRE I* and *DESIRE II* involved 20 partners from

8 European countries.[28]

One of the *DESIRE I* goals was to integrate all manually collected and quality-assessed collections of WWW resources with much larger robot-generated subject indices. The project was concerned with the analysis of subject search services based on selection, description and classification of high quality resources and regional search services based on the indexing of metadata generated by automated Web crawlers. The initial experience in manual (intellectual) classification was gathered, analysing three subject gateways that served as a test-bed: EELS, SOSIG and DutchESS. The project final report contained an analysis of different classification systems and it provided the first authoritative document that gave an extensive analysis of the use of classification on the Internet (Koch et al., 1997). In subsequent years this document brought the potential of library classification to the attention of the Internet community and was referred to in the literature by a number of authors (e.g. Newton, 2000; Cover, 2000; Chan, 2000 etc.).

DESIRE II (1997) aimed to develop an improved combination or a manually selected, catalogued and quality-assessed collection of WWW resources with a much larger robot generated subject index by providing cross-browsing and cross-searching. This was tested on the *Engineering Electronic Library System* (EELS) gateway (Ardö & Koch, 1999a).

WWLib - The Wolverhampton Web Library project from 1995 created a search engine and classified directory for UK resources with the option of subject browsing using the 100 top classes of the DDC. Like other classified directories, it offered the option of browsing the classification hierarchy or entering a query string. In subsequent years (to 1998) an automatic classifier was designed to compare the text on each Web page with entries in a 'DDC thesaurus' file. The thesaurus entries consisted of the DDC classmark and accompanying 'header' text. Gradually, this thesaurus was enriched with keywords and synonyms for each classmark required. As reported by the researchers, the overall success of the classification was closely dependent on the number of terms in the 'DDC thesaurus' and the creation and

---

[28] The ROADS project was completed successfully and it provided a toolkit for the maintenance and configuration of over 16 subject gateways. The toolkit consists of a software package of Perl modules that can be downloaded and installed on the service Web page. The toolkit provides programs for the maintenance and configuration of the gateway and includes an automatic link checker, statistical counters and a record validator.

enrichment of this thesaurus was described as laborious but necessary (Wallis & Burden, 1995; Jenkins et al., 1997).

The Renardus (http://www.Renardus.org) project was funded through the EU Information Society Technologies (IST) Programme. The project aimed to improve access to existing Internet-accessible collections of cultural and scientific resources across Europe by building upon the outcomes of *DESIRE II* (Renardus, 2000). *Renardus* built on the success of subject gateways based on a quality-controlled selection of resources aimed at academic and research communities. It proposed a distributed model where major subject gateway services across Europe could be searched together through a single interface provided by the *Renardus* broker. The project focused on devising models for sharing metadata, technical solutions, standardisation activities and developing business models to move from a pilot implementation to a fully functional service. The result of the project was a subject gateway based on mapping between different vocabularies and automatic classification of resources, which allows cross subject-gateway browsing (Koch & Day, 2001).

HILT - The High-Level Thesaurus was a UK funded project divided into two phases which looked into the use of a subject vocabulary across a range of communities and services (http://hilt.cdlr.strath.ac.uk/index1.html). *HILT I* was specifically interested in subject cross-searching and browsing and the conclusion of the project was in favour of a mapping between LCSH, UNESCO, UDC, Art and Architecture Thesaurus (AAT), and DDC vocabularies (Currier & Wake, 2001; Nicholson & Wake, 2003). In its second phase, it attempted to suggest subject vocabulary requirements at the collection level looking into the costs and benefits of a full terminology service.

3.3 Selection of classification for resource discovery

The success of subject access to electronic resources depends on the choice of a subject indexing language. Ideally, such a tool will enable the collocation of documents that have the same information content and navigation through a knowledge structure.

One of the benefits of the Internet is the degree to which resources can be made available to a broader audience. This can be enhanced through so called *alternative subject access*, i.e. additional subject orientations that make resources accessible to different audiences. Groups of users who do not share a common terminology or information behaviour can use the same service if this accommodates different indexing techniques and approaches. According to Hodge (2000) *alternative subject access* can be provided by: *(a)* indexing or classifying resources using multiple schemes from the outset, *(b)* adding new indexing terms while retaining original schemes and *(c)* mapping between the primary and alternative schemes. To facilitate discovery on this level, network technology has to support multilingual access and supply terms for expanding free-text searching. Library classification is a language independent indexing tool and is the natural choice to support various functions in resource discovery.

The existence of a classification scheme in more than one language may be considered as an important support to a multilingual search scenario. In this case, the classification serves as the central mapping vocabulary through which different verbal expressions are related (see Figure 3.1).

There are projects and ongoing research that focus on mapping thesauri and subject heading systems to provide access in cross-domain and multilingual environments, as well as developments in the area of multilingual thesauri and subject heading systems (Soergel, 1996; Doerr, 2001; Freyre & Naudi, 2003). These are, however, limited in subject coverage and difficult to create. The use of a general library classification for mapping different systems seems to give visible results quickly and easily (Hiom, 1998; Vizine-Goetz et al., 2004).

Because search terms are linked to a class number that comes with a semantic context, terms assigned to a classification scheme need not always be as rigorously controlled, as is the case with a thesaurus. This makes a classification system less specific but cheaper and easier to implement with the possibility of upgrading and expanding in any specific area. If specificity is required, it is possible to attach one or more thesauri or subject indexing systems to a classification scheme in order to achieve both a flexible and broad navigation structure and a precise, specific terminology for searching (Frâncu, 2000, 2002, 2003).

In summary, it could be said that the main idea of using classifications on the Internet is to provide a tool for controlled usage of natural language coupled with a subject navigation function, without undermining or excluding natural language.

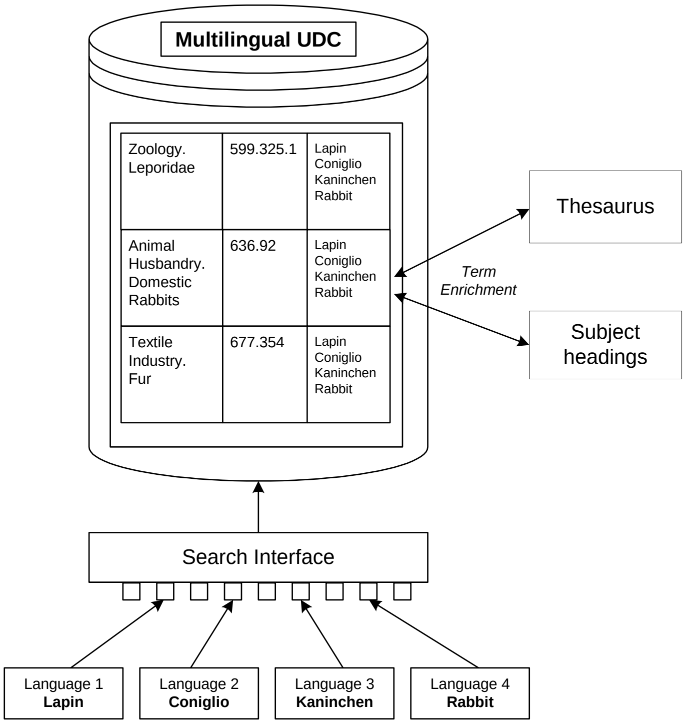


Figure 3.1: Simple model of using UDC in a multilingual environment

Classification can be successfully used in mapping different indexing languages or different domain languages. It helps to overcome ambiguities among variants of the same language and is useful whenever a search gateway has to bridge different knowledge areas, cultural or professional domains (Vizine-Goetz et al., 2004).

3.3.1 Evaluation issues

The Internet is not an environment on which a single indexing system can be imposed. The important characteristics of any indexing system used in this information environment are **extensibility**: to accommodate the growth of the vocabulary and subject areas, and **scalability:** to accommodate different degrees of depth in a subject area. The Internet can be used as a test-bed for every tool available in the documentation and bibliographic domain, starting from descriptive cataloguing, alphabetical and classificatory indexing languages, automatic clustering, automatic classification and automatic indexing (Thompson et al., 1997; Chan, 2000).

Classification structures used for knowledge organization on the Internet are of various kinds: some are home grown-systems of categories devised by general information services such as *Yahoo* or *AltaVista* and others are complex and stable library classification system tools (Koch, 1998).[29] Weinberg, described the former as an amateurish kind of classification with less significance to solving the problem of information organization and discovery. Besides being inconsistent in knowledge organization, as the category system grows, the problems and accusations of bias, lack of currency and lack of specificity will be encountered as was the case in the history of library classifications (Weinberg, 1996).

Because of the open nature of the Internet, there are many communities involved in information discovery. Different classification schemes, taxonomies and ontologies from businesses, industries, market sectors and products are deployed in resource discovery. Classifications like the North American Industry Classification System (NAICS), United Nations Standard Product and Services Classification (UNSPSC), Central Product Classification (CPC), Standard Industrial Classification (SIC) etc. contribute to the overall diversity of knowledge organization on the Internet (Cover, 2000).

Using a ready-made, stable, well-known classification scheme that has been tested in use and that covers either the whole of knowledge or some specialist subject area saves both time and resources. The significant theoretical background of classification systems, in particular with regard to facet analysis and complex subject synthesis, leads to compact classification schedules which allow very high specificity in indexing (Weinberg, 1996). In addition, internationally used general classification systems tend to be maintained and revised and are made available in an electronic form. These classifications hold the key to subject access for a large amount of data that are gradually merging to digital-born Internet collections. Classified catalogues to libraries, archive and museum collections are being linked to Internet resources and this enables easy transition between Internet searching and catalogue searching

[29] Lists of controlled vocabularies and/or classifications on the Internet are available at: http://www.lub.lu.se/metadata/subject -help (compiled by: T.Koch); http://www.public.iestate.edu/~CYBERSTACKS/ (compiled by G. McKiernan); http://orc.rsch.oclc.org:6109/classifiction/ (compiled by Scorpion contributors); http://sky.fit.qut.edu.au/~middletm/cont_voc.html (at Queensland University of Technology, compiled by Michael Middleton); http://www.fbi.fh-koeln.de/institut/labor/Bir/thesauri_new/indexen.htm compiled by the Laboratory for Bibliographic Information Retrieval) etc.

(Gödert, 2003).

The three internationally widespread general classification schemes that are often used on the Internet are: DDC, UDC and LCC.[30] The nature and the scope of the information service usually determine the choice and the level of sophistication of the classification that is applied. Thus, general search services tend to use universal classification systems, and subject specialized gateways choose special classifications. The majority of information services are, however, interested in the possibility of linking and combining different controlled vocabularies.

One part of the official *DESIRE I* report contained a study report called *The rôle of classification schemes in Internet resource description and discovery* (Koch et al., 1997) which offers a framework for an analysis of classification systems from the point of view of their application in resource discovery on the Internet. The following elements are found to be crucial:

- extent of usage in Internet services
- extent of usage in traditional and other online services
- multilingual capability
- strengths and weaknesses of the scheme
- integration between classification scheme and other controlled vocabularies
- digital availability
- copyright
- extensibility and development of the scheme

The *DESIRE* study was aimed at information specialists from different public domains. It was, therefore, valuable that it contained an analysis of large library classifications that are not usually used outside the bibliographic domain. Unfortunately the information provided in the report, although based on sound methodology, was inaccurate, misleading and patchy on the system comparison level, as is demonstrated in Table 3.1. The UDC, for instance, was obviously

[30] A distinction should be made between *internationally used general classification systems* (e.g. UDC, DDC, LCC) and *national* general classification systems (e.g. Nederland Basisclassificatie-BC; Sverige Allmama Biblioteksoforening -SAB) (Koch et al., 1997).

analysed by several authors not sufficiently informed about UDC or classifications in general and the comments in the final text appear poorly coordinated.

Table 3.1: Evaluation of UDC in DESIRE report

CRITERIA	DESIRE report	COMMENT
EXTENT OF USAGE IN INTERNET SERVICES	*...At least 5 gateways: BUBL, OMNI, NISS, SOSIG, GERHARD* *(BUBL in process to change to DDC* *OMNI in process of changing to NLM)*	It is not clear why the other simple subject directories using UDC that existed at time were not mentioned (see section 3.4).
EXTENT OF USAGE IN TRADITIONAL AND OTHER ONLINE SERVICES	*...There are no current figures available for the extent of usage but a survey in 1977 found that of thirty one countries in Europe, eight used UDC in their national bibliographies with a further four countries adopting the scheme in the 1980s (McIlwaine 1991)*	Some earlier sources mention up to 60 countries using UDC (Rigby, 1981; Batty, 1981). Bell mentions around 25 countries in the first edition of her book (1986) and 30 in the second edition (1997) (23 in Europe) - using UDC in their national bibliographies. According to 2004 research 112 countries world-wide use UDC (see Chapter Seven).
MULTILINGUAL CAPABILITY	*...The classification exists in several different languages including English, French, German and Japanese.*	Earlier sources mention up to 25 languages (Rigby, 1981). According to the research in 2004 (see Chapter Seven) the scheme is translated into at least 39 languages.
STRENGTHS AND WEAKNESSES	*...does not have good vocabulary in all subject areas*	It is not explained what is meant by 'good' or based on which edition of the UDC schedules this statement was made. At present there is no general classification for which it could be said that has 'good' vocabulary in all subject areas'.
	...main weakness of the scheme is that it is out of date"	Statement 'out of date' is not true for the UDC as a whole after 1990.
	...complex structure of the scheme is considered to be a problem [because] main tables of the classification can be combined with auxiliary tables.	Organization of vocabulary into main and auxiliary schedules is a normal feature in majority of library classifications.
	...SOSIG and NISS suggested that UDC is not updated often enough: some subjects are outdated, weak in environmental studies and developments studies	SOSIG and NISS did not buy any of new version of the UDC from 1990s and their comments are based on the 20 years old editions of the UDC for which they do not have to pay for a user licence.
	NISS finds it too complicated to use composite classification codes and decimal notation does not reflect a true hierarchy	It is not necessary to use complex UDC numbers. No existing library classification is known to reflect fully the 'true' hierarchy using notation only.
INTEGRATION BETWEEN CLASSIFICATION SCHEME AND OTHER CONTROLLED VOCABULARIES	*"The ETHICS (ETH library Information Control System) at ETH (Eidgenössische Technische Hochschule) in Zürich link each UDC number to related thesaurus terms in English, French or German."*	UDC is linked to subject headings systems or different controlled vocabularies in many library catalogues world-wide. ETHICS is only exceptional because of it multilingual authority file.
DIGITAL AVAILABILITY	*UDC MRF file created in 1993 and is currently maintained by Technical Director in the Hague*	Comment on maintenance is both wrong and irrelevant while important information about annual update, export formats is lacking.
COPYRIGHT	*"copyright of MRF belongs to UDC Consortium and can be purchased"*	Correct but contradictory with the rest of text.
	"It is usually free to use. None of the services (except GERHARD) had had paid for to use UDC, and had no problems in over copyright when they published sections of it on their Web pages (there would have been copyright or licensing issues had the services used the UDC MRF)	Old, printed UDC schedules are not protected by UDCC copyright licence, which is why they were used by NISS and SOSIG and which is precisely why these services complain about UDC being obsolete in the section above
EXTENSIBILITY AND DEVELOPMENT OF SCHEME	*'not maintained often enough'*	Since 1993, UDC MRF is updated and distributed as a new version annually.

In spite of the weaknesses mentioned here, the *DESIRE* report played an important rôle in highlighting real issues that are of interest to information services on the Internet. It also pointed out the obvious problems in finding basic information about, for instance, UDC that would be readily available for those outside the library domain and is not scattered across textbooks and research articles.

Since 1993 to January 2005 there was no publicly available information on licence pricing for use and for publishing and about the amount of time necessary to obtain such a licence. An outline of UDC (around 350 numbers) is, for instance, published on the Web by the UDC Consortium and nowhere does it state whether or not this is free to use and under what conditions. Classificatory structures that are free to use have great value for Internet use. The lack of recent data on UDC usage figures; for instance, the lack of clear and more transparent information about UDC export formats and pricing, may also be an issue.

The fact that Internet users are not well informed about the UDC appeared, again, evident in the HILT report (Currier &Wake, 2001: 10) which summarised the strengths and weaknesses of UDC as shown in the Table 3.2. Confused criteria and different levels of information for individual indexing languages become, again, more visible when information on similar indexing tools is compared and quality mentioned for one system is simply ignored when evaluating another.

Table 3.2: Strength and weaknesses of UDC according to HILT project

STRENGTHS	WEAKNESSES
• conforms to international thesaurus standards • reasonably well resourced and supported • available in several European languages • browsable on web • simplicity of use for indexing high level descriptions • wide subject applicability • detailed analysis of relationships between terms	• too general, not specific enough • variable development of subject areas • does not accept amendments or addition suggestions

Apart from the weaknesses in the HILT research methodology when comparing subject alphabetical indexing languages and classificatory indexing languages, a number of statements in Table 3.2 are inexplicable, such as the fact that *'conforms to international thesaurus standards*' or is *'browsable on the web'*, *'simplicity of use for indexing'* or *'does not accept amendments or addition suggestions'*. Some statements are contradictory such as declaring *'detailed analysis of relationships between terms*' a strength and *'too general, not specific enough'* as a weakness. But only when one compares this to DDC or other systems does the lack of expertise of the report's authors become more obvious.

3.4 UDC in quality subject gateways

The first subject gateways using library classification were created by individuals and enthusiasts as a part of free and uncoordinated efforts towards the organization of Internet resources. They tried to compensate for the inadequacy of search engines in collocating subjects. Subject gateways using library classification introduced by academic and research projects were available as early as 1993 and continued to appear throughout the 1990s. These free available information services were funded by national or international projects and they encompassed the following kind of applications of library classifications:

- clerical classification of links on small to medium-size gateways, with or without the help of simple metadata
- simple classification of a large number of harvested resources using harvesting and metadata creation tools and more advanced technology
- automatic harvesting, indexing and classification

Table 3.3: Overview of quality subject gateways using UDC 1993-2005

	CREATED	STATUS 2004	SUBJECT COVERAGE	LANGUAGE	INDEXING
1. WAIS/WWW	1993	not updated	general	English	automatic
2. BUBL	1994	operational, no UDC -1996	general and LIS	English	manual
3. OMNI	1995	operational, no UDC -1998	medicine	English	manual
4. NISS	1995	operational	general	English	manual
5. SOSIG	1994	operational	social sciences	English	manual
6. FVL	1996	operational	general	En., Fin., Swed.	manual
7. GERHARD	1998	operational	general	En., Fr., Germ.	automatic
8. OKO	2000	operational	general	Eng., Sloven.	manual
9. PORT	2000	operational	maritime information	English	manual

Bulletin Board for Libraries (BUBL) was the first quality subject gateway in the U.K. that used UDC and it set an example for other services to follow. During the period 1990-2004, UDC was frequently mentioned in relation to the following nine, medium and large, quality information services: Nordic WAIS/WWW, BUBL, FVL, OMNI, NISS, SOSIG, GERHARD, OKO - Slovenian catalogue of the Web resources, PORT. Some, such as BUBL and *Organising Medical Networked*

Information (OMNI) are no longer using UDC.[31] Smaller directories such as IANUS - *Informazioni Classificate per Discipline*,[32] *Services in classified order* at the University of Wales, Aberystwyth[33] and *Plambeck UDC directory*,[34] ceased to exist.

After 2000 many new subject portals and gateways appeared in Central and Eastern Europe, especially in the Russian Federation. Some of these gateways may be using UDC for simple subject browsing such as *EJOL - Electronic Journals Online Library* (http://ejol.irb.hr/). More important are quality subject gateways using UDC for searching, browsing and vocabulary mapping such as the Russian *Federal Education Portal* (http://www.edu.ru) or the Czech *UIG - The Uniform Information Gateway* (http://www.jib.cz).[35] Because they do not provide an English search interface they are not going to be described here.

In this chapter a brief description of the seven operating services (five manual and two automatic) will be given. At the end of this section a summary analysis of functionality for all will be provided.

3.4.1 Manually indexed subject gateways using UDC

All five gateways that are described below can be considered large or medium scale quality information services. They use metadata and both indexing and metadata production are manual. All of them are freely-available Web-based services.

NISS - National Information Services and System at http://www.niss.ac.uk (Bath University) has been maintaining professional information services for UK education since 1988. In 1990s, its function was extended to government information, health and other sectors. One part of the NISS service using UDC is the Directory of Internet Resources (Lafford & Stone, 1997).

The NISS system is modelled on library practice i.e. it is based on a catalogue

---

[31] BUBL changed from UDC to DDC in 1996 as the service obtained permission to use DDC from OCLC. OMNI changed to the NLM classification in 1998 as it is more suitable for resources in the field of medicine and is freely available.
[32] Previously at http://www.ianus.cineca.it/venus/ionio/www/awww.htm.
[33] Previously at http://www.aber.ac.uk/~infolib/classified.html.
[34] Previously at http://www.chem.ualberta.ca/~plambeck/udc/index.htm.
[35] It is planned to connect to UIG subject gateways using DDC (such as Renardus) through UDC-DDC concordance tables as a bridge with approximately 500 classes (Stoklasová, Baliková & Celbová, 2003).

of Internet resources and uses simplified UDC numbers to provide subject searching and browsing. Up to 2003, the subject gateway was browsable by inverted 'subject tree' UDC menus, by UDC classmark, and by alphabetical subject heading (see interface screen capture in Appendix 2.1, Figure 2.1). Since 2003, the UDC subject hierarchy has not been available for browsing (see existing subject categories in Appendix 2.1, Figure 2.2) and UDC numbers can only be seen in resource metadata.

SOSIG - Social Science Information Gateway at http://www.sosig.ac.uk **(**Bristol University**)** is an Internet catalogue containing around 10,000 quality selected information resources for students, academics, researchers and practitioners in the social sciences, business and law. Launched in 1994 as an information service for social sciences it was re-launched as a social science, business and law hub with a new interface and additional features[36] in 2000 (Huxley, 2000). SOSIG was one of the most important gateways used as a test bed for the *DESIRE* project and it has applied the project's automatic Web harvesting technology. Since then, SOSIG has continued exploring XML/RDF technology in the storage, query and interchange of controlled vocabularies such as HASSET and UDC (Hiom, 1998).

At the beginning, SOSIG used a total of 161 UDC numbers. A smaller selection of these classes was made available for browsing and the remainder was used for cataloguing. The detail of the numbers used to vary from being at the top of a hierarchy (e.g. Philosophy = 1) to being fairly specific (e.g. Environmental Issues = 551.588 or 658.562 = Quality management). Interdisciplinary resources or those treating more than one subject were given as many UDC numbers as necessary in a manner of post-coordinate indexing (Hiom, 1998). The UDC browsing hierarchy (Appendix 2.2, Figure 2.3) was replaced by main subject sections with 500 sub-sections (Appendix 2.2, Figure 2.4). UDC is still used in the process of indexing to control keyword assignments for 13 out of 16 main subject areas (SOSIG Guidelines for assigning keywords, 2004).

Catalogue OKO at http://www.zrc-sazu.si/oko/ is a Slovenian catalogue of Web resources hosted at *The Scientific Research Centre of Slovenian Academy of Sciences and Arts (SRC SASA)*. The catalogue covers quality resources on the Web

[36] These include personalized alerting services Grapevine and MyAccount, which provide a selection of resources, conferences, courses and communication channels for selected areas.

irrespective of the language or the source providing the resources are judged trusted and useful for the Slovenian academic and research community.

Resources are classified manually using UDC and can be searched and browsed in English and Slovenian (an alphabetical subject sequence and a UDC sequence are available for browsing). The UDC hierarchy contains 2-3 levels, while synthesised UDC numbers are searchable only as whole strings (Appendix 2.3). The UDC edition used is from the 1980s (with the exception of *004 Computer science*). The total number of UDC classes used is around 700.

FVL - Finnish Virtual Library (in December 2004 changed to *Science Linkhouse*) at http://www.linkkitalo.fi/ is an outcome of the *Finnish Virtual Library Project* initiated in 1996. It is a quality subject gateway with general knowledge coverage that provides searching and browsing of Internet resources. The main language of the service is Finnish with a large number of resources in English (one database is in Swedish). The FVL cooperated with the *Renardus* project and is now accessible through the *Renardus* service. The FVL i.e. Science Linkhouse combines its own classification with UDC, DDC and NLM Classification. UDC is used as an underlying structure for extractions of subject classes for browsing in certain areas and does not show the UDC hierarchy or numbers on the interface (Appendix 2.4).

PORT - Maritime Information Gateway is the National Maritime Museum's subject gateway at http://www.port.nmm.ac.uk/about.html that provides catalogues of around 2500 Internet resources containing maritime information. The gateway is created in collaboration with ILRT (Institute for Learning and Research Technology at the University of Bristol) and is based on the model developed by other UK subject-based information gateways as part of the Electronic Libraries Programme. Librarians in the National Maritime Museum select resources, classify them manually using UDC and create metadata.

Resources are organized into twenty subject categories which, at first glance, have no resemblance to the UDC structure (Appendix 2.5, Figure 2.7). The display of a full subject hierarchy (listed alphabetically by main heading), however, shows a structure that is obviously based on more complex UDC numbers (Appendix 2.5, Figure 2.8). Each subject category can then be browsed choosing period, form, place or some other helpful facet and it is obvious that this kind of restructuring and reorganization of display is achieved with a help of the UDC (Ashton & Robertson, 2000).

3.4.2 UDC gateways with automatic indexing

Automatic text processing and indexing is usually implemented on textual resources in a single language. When index terms are, however, mapped to classification vocabulary, resources may be searched in any language in which the classification is made available. Between 1993 and 1999 there were a number of projects that explored the application of library classification for automatic indexing. Research was focused on the ability of general library classifications to group and organize resources into smaller subject-related "chunks".

One of the research institutions that started to explore this field early in 1992 was the NetLab - the Research and Development Department at Lund University Libraries (http://www.lub.lu.se/netlab/). Members of the research team participated in several important projects on automatic indexing of Web resources using classification including, *Nordic WAIS/WWW* project, *DESIRE I*, *DESIRE II* and *Renardus*. OCLC's *Scorpion* project, mentioned earlier, was one of the most important contributors to the field of automatic indexing and classification of Web resources. One outcome of this research project is the *Scorpion Open Source* software that is made available for everyone who has machine-readable classifications or thesauri and would like to test the software for automatic indexing (http://www.oclc.org/research/software/scorpion/).

The approach explored in the *DESIRE II* project, for instance, consisted of harvesting pages combining different methods in following links from a small number of high quality manually selected engineering resources. The term index generated from the harvested pages was then matched to the *Engineering Information thesaurus* (Ei) containing 16,000 terms organized in 700 Ei classes. Term weighting was calculated according to the term complexity, match location and matching frequency. The evaluation showed an excellent result of 100 percent correct classification and comparison with manual indexing showed 57 to 66 per cent agreement (Ardö et al., 2000).

DESIRE II demonstrated that the use of a library classification assists in achieving good results in automatic indexing with very simple weighting algorithms and simple heuristics (Ardö & Koch, 1999a). Matching to the Ei thesaurus was then compared to automatic indexing of engineering resources based on DDC using the Scorpion program at OCLC. Further research was planned for improvements using linguistic software (noun phrase extractors and morphological extractors) and heuristics, in co-operation with the *GERman Harvest Automated Retrieval and Directory* (GERHARD) project as well as the comparison of the performance with *Scorpion*. The research was originally intended for exploring the possibility of expanding and connecting the vocabulary between these systems (using three different classifications Ei, UDC and DDC) (Koch & Vizine-Goetz, 1998). The outcomes and experience from *DESIRE II* were taken further in the *Renardus* project (Koch, Neuroth & Day, 2003).

UDC was the first classification used for automatic indexing of Internet resources and the only one that supported automatic indexing of Internet resources with a multilingual search interface. There were two quality subject gateways using UDC in the period 1993-2005.

Nordic WAIS/WWW, at http://www.lub.lu.se/auto_new/UDC.html, is one of the first Internet information services to use classification and was certainly the first to apply automatic classification in 1993. The project was funded by *Nor Info* and involved the *Danish Technical Library* and *the Lund University Library*. The project goal was to apply a library system model on the Internet and to build a gateway for the Wide Area Information Server (WAIS) databases. Its primary aims were to enable automated processing of WAIS source files to produce classified indexes available over the WWW, and to develop better gateways between WAIS and the WWW. In its final report it was established that the project successfully demonstrated the possibility of network-based resource discovery and retrieval through a subject gateway based on library classification. (Ardö et al., 1994).

The project had three parts: demonstration of how an existing library system (ALIS) can be integrated in open client/server architecture, simplifying the use of the WAIS database by establishing a WWW interface and combining hypertext browsing with a search option (Appendix 2.6). The project used only 51 UDC numbers to classify approximately 700 WAIS database descriptions which were quite homogeneous in their content.

Vocabulary from the WAIS databases' description was indexed and compared to UDC vocabulary using only two top levels of division (there was no explanation as to whether or not it was only from the UDC caption or whether a richer index was created for the purpose). The classification number was automatically assigned to the source when a term match was established. Based on the final classification, the individual WAIS database was linked into the gopher WAIS UDC subject tree. The classification was also used to build a hypertext hierarchy corresponding to the UDC structure. Nordic WAIS/WWW is not updated as the WAIS databases are no longer published.

GERHARD - German Harvest Automated Retrieval and Directory, at http://www.gerhard.de, is a fully automatic indexing and classification system of academic resources on the German Web and was developed as a research project at the University of Oldenburg. The project had two phases (1997 to 1998 and 2001-2002) and was funded by the *Deutsche Forschungsgemeinschaft*. *GERHARD* is a database-driven robot that collects academically relevant documents which are indexed using computer-linguistic and statistical methods and classified by UDC. The user interface is trilingual (German, English and French) and allows the user to look for similar documents very easily through its tight integration of searching and browsing mechanisms. In the second phase of the project statistical methods were improved, the UDC vocabulary was extended and the database was enriched with some manually classified resources.

GERHARD's architecture consists of a database-driven gatherer, fast automatic classification and an integrated searching and browsing service (Appendix 2.7). The generated metadata and index of documents are held in a relational database (ORACLE) which now contains almost 1,300,000 records. Automatic classification is based on the UDC authority files from the ETH library system in Zürich (a system called ETHICS[37]). The ETHICS subject authority files support searching of compound and complex UDC numbers using indexing terms (in English, German and French) or numbers and it supports about 15 different relationships that could be established between UDC numbers. It is a reliable controlled vocabulary which consists of approximately 70,000 entries.

The first step in *GERHARD* was to extract a vocabulary called UDCZ-Lexicon from the ETHICS subject authority file. The creation of UDCZ-Lexicon was done automatically by conversion of 500,000 lines of text from ETHICS to language expressions that normally occur on Web pages. Conversion consisted of

---

[37] The ETHICS Library Information Control System was developed and used in the ETH library from 1983-1999. In 1999 ETH migrated to the Aleph vendor system to support NEBIS (*Netzwerk von Bibliotheken und Informationsstellen in der Schweiz*) that consists of over 70 academic libraries (Pika, 2002).

morphological analysis of each word in the UDC entry and its reduction to a stem. To each stem a variable was added in order to enable a specific form of words that may occur in the text. Once made, this lexicon was compiled to serve as a “recogniser” with the purpose of identifying regular expressions and providing the corresponding UDC notation.

German Web pages were harvested into the database and underwent text analysis to extract words that were matched to the UDCZ-lexicon and then assigned a notation. Notations were clustered and an occurrence frequency was weighed and statistically compared with the UDCZ database. Additionally, quality and specificity were checked using special algorithms until the average number of 14 per document was reduced to the best six-eight notations (secured by threshold applied). Finally, there was a statistical post-processing in which doublets of documents were weeded by matching the title of the document with titles already in the database. Titles are analysed and matched to exclusion lists produced to prevent classifying pages such as e.g. Web server statistics. The notation gathered from titles, headings and body text are matched and their concordance and precision is further checked through special algorithms.

GERHARD permits both navigation and searching through a subject directory. Each subject class is hierarchically displayed and is represented with a short class caption. Moving through the hierarchy of categories is easy and user-friendly. Against each class there is a 'hyperlinked' number of documents available and users can browse the list of documents for each level of the hierarchy (http://www.gerhard.de). Within each class documents are ranked according to accuracy. This is said to be achieved through the use of the complex UDCZ/SOIF (Summary Object Interchange Format) design, (Möller et al., 1999). *GERHARD* is the most sophisticated application of UDC, and classification in general, for the purpose of the Internet resource discovery.

3.4.3 Summary on UDC subject gateways

Up to the year 2000, UDC had a prominent rôle in resource discovery (cf. Newton, 2000). Table 3.3, at the beginning of section 3.4 shows a list of nine medium to large size quality subject gateways using UDC in the period 1993-2005 with an English language interface, two of which are based on automatic indexing/classification. In section 3.4.1 and 3.4.2, seven services that still use UDC are described. Three non-UK based gateways have an interface in languages other than English. Only two of these gateways emerged after 2000 (OKO and PORT) and only two are special in subject coverage (PORT and SOSIG).

Table 3.4, shows a summary overview of interface functionality that can be observed for seven operating subject gateways using UDC. Four of these systems use very little of the UDC vocabulary (NISS, OKO, SOSIG, WAIS/WWW) and three do not use UDC synthesised concepts. NISS and SOSIG (and BUBL and OMNI before they changed to other classifications) used old printed editions of the UDC in order to avoid licence costs. Only GERHARD paid the UDC MRF licence.

Two gateways PORT and FVL (now Science Linkhouse) are probably using some of the recent editions of the UDC but it is not possible to confirm this assumption, as they are not displaying the UDC structure on the interface or UDC numbers in the metadata. However, they contain subjects that are pre-combined concepts in UDC. All services are based on metadata with no UDC notations shown. The general trend is, also, not to display UDC numbers on the interface (only WAIS, which is not updated, and OKO do). In 2004 only three services display a UDC hierarchy and show UDC captions compared to five before 2003. FVL, NISS, PORT and SOSIG have a gateway specific subject directory for browsing.

Table 3.4: Gateway interface functionality based on UDC

	Hierarchy[38] display	Hierarchy level[39]	caption	number synthesis	shown in metadata	number search	search to browse	Use UDC MRF
1. FVL	NO	---	NO	YES	NO	NO	YES	----
2. GERHARD	YES	no limit	YES	YES	NO	NO	YES	YES
3. NISS	2001> NO	---	2001> NO	NO	YES	NO	NO	NO
4. OKO	YES	1-4	YES	YES	NO	YES	YES	NO
5. PORT	NO	---	NO	YES	NO	NO	YES	---
6. SOSIG	2000>NO	---	2000> NO	NO	NO	NO	NO	NO
7.WAIS/WWW	YES	3	YES	NO	NO	NO	NO	NO

Gateways show a very different approach to exploiting classification and it is obvious that user friendliness is achieved by hiding not only the UDC notation but

[38] Subject hierarchy shown on the interface UDC (with or without UDC numbers?).
[39] Approximate number of hierarchical levels of subjects available for UDC browsing.

also by altering the subject browsing structure and captions. None of the subject gateways (apart from OKO) allows a search of UDC numbers. SOSIG and NISS have replaced the UDC browsing structure with a list of subject headings. Subject headings in NISS, SOSIG and PORT represent a shortcut to the subject areas of interest to the majority of users and reduce the length of navigation in order to present the subject areas that are most sought. While it was easier to describe the level of UDC hierarchy applied in the early phases of general subject gateways (e.g. WAIS/WWW, NISS) this is not so with subject orientated subject gateways or those that use classification in the background. Gateways seem to move more towards using classification as a subject framework for the purpose of semantic linking of keywords, vocabulary control, help in indexing, and mapping subjects between different vocabularies, different languages and different collections.

After 2000, the use of the UDC in subject gateways in the English-speaking part of the Internet decreased from 9 to 7. There is a slight difference if the East European countries are taken into account. The research done by Stoklasová shows that the UDC is the predominant library classification in subject gateways in the Russian Federation and Czech Republic (Stoklasová, 2003).

A comparison between DDC and UDC as reported by the *Renardus* project in Table 3.5 clearly indicates how the previous research and testing of classification system can lead to the 'enhancement of classification data' and can clearly influence and determine the future use of a particular scheme and selection of classification for cross-browsing/searching. According to Neuroth & Koch (2003) the advantage of the DDC over the UDC is the fact that DDC has 'much larger user community', 'research based development', is mapped to LCSH and is used in many gateways.

The recent report that included research on subject gateways in Australia, Czech Republic, Norway, Russia, Sweden and USA given by Stoklasová (2003) shows that the predominant classification in these countries was DDC. With UDC more frequently used only in information services in the Czech Republic and Russia, the author emphasizes the importance of mapping different classification systems to DDC. The research on DDC and its presence in subject gateways made a strong case for its use in *Renardus* (Table 3.5), and, in return, the DDC presence in *Renardus* has influenced its further use in other national and international gateways that plan to link to Renardus.

Table 3.5: Comparison between DDC and UDC in selecting the classification

CRITERION	DDC/UDC COMPARISON (Neuroth & Koch, 2003)
ONLINE AVAILABILITY AND TOOLS	The DDC was adapted earlier and more quickly to usage in digital systems via the Internet. It is completely and easily available as "Web Dewey" for all Web browsers and platforms. Furthermore the DDC is connected to large databases of real documents via the Web, incl. the CORC service, that allow the verification of potential subject content of classes and the correct usage of the classification.[40]
GLOBAL USAGE	The DDC has a much larger user community on a global level. DDC is more frequently chosen as the classification system in Internet resource discovery services. Some gateways already use DDC or provide a mapping to DDC.
SUITABILITY OF THE CLASSIFICATION SYSTEM AND ITS FUNCTIONALITY	The UDC is probably more "modern" and has made faster progress towards a faceted structure. Its way of allowing the construction of new special classes by a straightforward relationship between existing ones is in the Renardus application context, however, a disadvantage. In addition, the UDC has 61 000 classes; three times larger than the DDC and might be far too detailed for Renardus' purpose.
FREQUENCY AND CHARACTER OF THE UPDATES:	Both systems have large amounts of outdated captions and class structures. Both are slow with integration of external standard vocabularies.
RESEARCH AND METHODOLOGICAL DEVELOPMENT EFFORTS	With the rich, broad and long-term activities of the OCLC Office of Research Knowledge Organization Group and related Dublin Core metadata and XML/RDF work, the DDC has a decisive advantage when it comes to research based development potential and is future proof in digital library contexts. So the enhanced DDC contains intellectually and statistically mapped vocabularies like the LCSH which are extremely useful in classification and mapping work.
CONCLUSION	The main advantage of DDC here is the effort that has been made to adapt captions of the system to general end-users and, even more importantly to Renardus, to the real content of digital documents in today's Internet.

UDC was more popular in the enthusiastic research community in the earlier phases of Internet development (up to 1995) when a selection of UDC numbers from old printed editions were used, for which there were no restrictions on use in classifying documents. The research in improving classification data, by enhancing the subject alphabetical index and testing its functionality were, for instance, encouraged by the owners of LCC and DDC, who allowed the systems to be used in research without cost, providing they had access to the research results. This was not the case behind the policy of the UDC owners in 1993-2005 with its business entirely focused on selling UDC as 'raw data' or half product, i.e. schedules without a subject-alphabetical index to publishers or libraries who can afford to pay for both UDC and its redesign as a functional indexing tool. Research on the improvement of UDC data for use in a networked environment is, obviously, far from being encouraged in this case. Permission to get UDC data for research purposes may take months or sometimes a year to obtain while conditions and costs are not known to the project

[40] The fact that it is considered to be more rapidly adapted to the Internet through Web Dewey may not be a valid argument as UDC-online was made available to users earlier than DDC, while the UDC database has been distributed since 1993.

bidder and cannot be planned in advance.[41]

The presence of library classification in quality search gateways and portals seems to be as equally relevant in 2004 as it was in 1994. The infrastructure behind browsing directories, however, has significantly changed and grown in complexity and classifications are gradually moving from their 'naive' presentation on the interface to metadata and authority files where they are used to control natural language terms. For this purpose a subject vocabulary is often managed as a database tool and is shared through a separate repository.

Very few directories are still based on manual indexing and are not coupled with powerful search engines. Gateway interfaces are greatly improved by enabling browsing using interactive frames and navigation through hyperlinks that span windows and frames and allow the display and easy tracking of browsing steps. Traditional classification structures have been replaced by a user-orientated organization of resources.

Since the year 2000, subject portal taxonomic classification structures have tended to be combined with, or replaced by, a faceted organization of object/subject properties and attributes (Merholz, 2001; Devadason, 2003; La Barre, 2004; Ramshirish, 2004). The infrastructure for a faceted vocabulary with tools and a vocabulary format is being created and tested for use for this specific purpose (Van Dijk, 2003).[42]

The popularity of faceted classificatory as opposed taxonomical structures simply demonstrates that a new window of opportunity is opening for UDC, providing its data is made more affordable and available in formats that support facet browsing. While UDC vocabulary may be considered to be too large and specific for gateways such as *Renardus*, this could be an advantage for automatic indexing using classification.

The expansion of the UDC vocabulary through its use in a library catalogue (mapped to descriptors in several languages) as exploited in GERHARD seems to be very relevant. The fact that at present the UDC MRF is not being integrated with any

---

[41] The first publicly available information about UDC MRF licence price was published in December 2004 in the *Extensions & Corrections to the UDC* 26 (Appendix 6.4).

[42] In parallel, the report on user behaviour based on log analysis of the Renardus service, which offers browsing by DDC, shows strong predominance (80%) of a subject browsing function (Koch & Ardo & Golub, 2004).

external vocabulary and is not being tested in any research project on resource discovery, however, may have an impact on its future.

3.5 Classification 'schemes' in metadata 'schemas'

Metadata is the basic infrastructure for Internet information services and the actual 'carrier' of classification data, when these are used in quality subject gateways. Hence, the limitations and potential of metadata standards and general trends in their development are very important for the use of controlled vocabulary.

Table 3.6 shows five metadata schemas that are examined in this section: DCMES, LOM, EAD, AMICO Data Dictionary and MARC.[43] The objective is to evaluate the ability of given metadata formats[44] to accommodate subject description. The approach is to examine the choice of data elements that may hold subject description (classification in particular), the ability of metadata syntax to declare, identify and code classification data and the way metadata instances are linked to an external vocabulary scheme i.e. authority file.

Table 3.6: General characteristics of metadata schemas

	DCMES	LOM	AMICO	EAD	MARC
BAND[45]	SIMPLE	COMPLEX	COMPLEX	COMPLEX	COMPLEX
DOMAIN	GENERAL	EDUCATION	MUSEUMS	ARCHIVES	SPECIFIC: LIBRARY
PRODUCTION	AUTOMATIC/ MANUAL	MANUAL	MANUAL	MANUAL	MANUAL
EMBEDDED/ STANDALONE	EMBEDDED; STANDALONE	STANDALONE	STANDALONE	STANDALONE	STANDALONE
ENCODING	HTML, XML, RDF	XML	XML	SGML, XML	MARC FORMATS, XML
VOCABULARY[46]	YES	YES	YES	YES	YES

Selected metadata standards share characteristics that are typical for the majority of metadata standards developed to support resource discovery in the networked environment (cf. Taylor A. G. 1999, 1999a; Taylor, C., 1999; Caplan,

[43] MARC, as the most complex of metadata standards, will be described in more detail in Chapter Six, and is presented here for the purpose of comparison.

[44] All standards shown here apart from MARC, which has various formats, and DCMES, have a single format or schema.

[45] To which band a metadata standard belongs: simple structured generic formats or complex metadata format.

[46] Does the metadata standard recommend some existing standard subject indexing language such as classifications, thesauri or subject heading systems?

2003). Most of these metadata standards have a complex structure, metadata is as created and maintained as standalone and is encoded using XML. Concerned with interoperability, and as a matter of good practice, most of these standards recommend the use of standard controlled vocabularies, such as classifications, thesauri and subject heading systems.

Metadata schemas tend to separate content description into smaller structured units. The complexity of subject description leads to the complexity and diversity of descriptive elements that accommodate this kind of information. Because of this, and when planning to use classification in metadata, the following facts may be relevant:

- subject description may be 'split' between several metadata elements
- classification schemes, such as UDC, may be used in metadata elements other than subject (e.g. physical description, audience)
- a classification number represents a complex statement even when the subject it describes is simple (a classmark contains information about a hierarchy context, has a description and can contain cross references to other subjects)
- not everything that may be defined as a subject can be satisfactorily indexed by a classification such as UDC (e.g. a subject can be another resource, or the creator)

The first point, on 'subject scattering', can be illustrated with the bibliographic metadata model as shown in Figure 3.2 (based on Functional requirements for bibliographic records, 1998:15). Subject description is a specific group of entities that consists of concept, object, event and place.

The second point refers to the fact that a large classification, such as UDC contains many useful smaller vocabularies that can be used independently. These may include facets such as languages, persons, materials, forms/document types and places, which can be useful in descriptive elements, such as *coverage*, *type*, *audience,* and not only in describing a subject in its narrow sense.

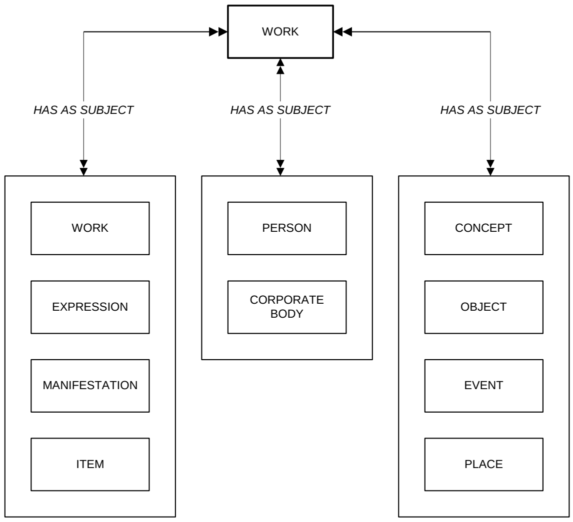


Figure 3.2: Subject relationships according to the FRBR

Subject description is almost always a duplication or summary of information about content that may be partly present elsewhere in the description (description, abstract, coverage, title, keywords). This redundancy is an expected and purposeful duplication that has its roots in the complexity of information systems and functions as a contingency in IR and information exchange.

3.5.1 Qualified Dublin Core Metadata Element Set (DCMES)

DCMES was created initially to serve as a simple and generally applicable resource description that could be embedded in a document's HTML source, which is why it is designed as a generic standard for simple resource description, applicable in a wide range of areas (Weibel, 1995; Dempsey & Heery, 1998). Currently, DCMES is the most frequently used description standard in quality subject gateways and it underpins many information services across different domains. It is used both as HTML embedded metadata as well as a standalone resource description, usually encoded as XML or XML/RDF. Since its creation in 1995 DCMES has matured in terms of its application scope, structural and semantic refinements. These developments are guided by an international community of users and implementers that are represented by the Dublin Core Metadata Initiative (DCMI) - http://www.dublincore.org.

Structural characteristics. DCMES has a simple flat structure of 15 descriptive elements with intuitive element names, functions and semantics. The choice, number, order and occurrence of the elements is arbitrary. Practical DCMES applications created the need to refine and qualify metadata elements as these were not specific enough. This initiated the development of two types of Dublin Core Qualifiers that were supposed to meet the requirements of further semantic specificity: **refinements** and **encoding schemes**. While refinements help to specify the element (e.g. the refinement for 'title' is 'alternative'), the **encoding scheme** identifies a value that is going to be used in a metadata element. The encoding schemes can be controlled vocabularies, some other formalised notation or machine-processable data: ISO standards, classification schemes, thesauri etc. The introduction of encoding schemes in DCMES was a necessary requirement for the use of knowledge organization tools such as library classification. Once library classifications were recommended for use in the DCMES this became a model for many other metadata standards and their schema to follow.

Three DCMES elements are used to describe the subject content of a resource: *dc:subject*, *dc:description* and *dc*:*coverage*. While *dc:description* contains unstructured text (abstract, summary) of the content of the resource, *dc:subject* and *dc:coverage* are suitable to hold the classification number. Element *dc:subject* states what the resource is about (topic, event, person, time, place). If the resource is **about** some temporal or spatial concepts (time and space) this may be encoded in *dc:coverage*. In the Qualified DCMES recommendation from 2001 the suggested vocabularies (encoded schemes) to be used for spatial data are DCMI Point, ISO 3166, DCMI Box, TGN -Thesaurus of Geographic Names, and for temporal data DCMI Period and W3C-DTF. The choice of encoding scheme is, however, completely up to the implementers and *temporal* and *spatial coverage* could be also expressed using concepts from classification schemes if this may be considered advantageous i.e. if some hierarchical presentation is necessary. Subject element is repeatable and can be populated with:

- keywords
- terms from standard vocabulary scheme (LCSH, DDC, UDC etc.)
- using both keywords and one or more vocabularies

When a standard vocabulary scheme is used this has to be identified through the qualifier 'scheme' and the classification number is then entered as a value in the form of a simple text string, as shown in the following example:[47]

```
<dc:subject content="databases, database management systems, DBMS">
<dc:subject scheme="UDC" content="004.65">
```

As DCMES is used in different domains, it will tend to be populated with different vocabularies. This is facilitated by the use of an XML namespace syntax for metadata transport which allows the mixing of vocabularies from different metadata standards. XML/RDF, for instance, will support the description of an object with descriptive elements being taken from different metadata schemas and will allow the encoding of any external vocabulary. The DCMES schema is now made available (from October 2002) in a stricter XML/RDF format. In this way, each of the DCMES elements, their qualifiers (encoding schemes and refinements) are assigned a *Uniform Resource Identifier* (URI) within its declared DCMES 1.1 namespace. When UDC is used in DCMES it can be declared as element refinements as shown below:

```
URI:http://purl.org/dc/terms/UDC
Namespace:http://purl.org/dc/terms/
Name:UDC
Label:UDC
Definition:Universal Decimal Classification
See:http://www.udcc.org/
Type of term:http://dublincore.org/usage/documents/principles/#encoding-scheme
Qualifies:http://purl.org/dc/elements/1.1/subject
Status:http://dublincore.org/usage/documents/process/#registered
Date issued:2000-07-11
Date modified:2002-06-15
Decision:http://dublincore.org/usage/decisions/#Decision-2002-01
This version:http://dublincore.org/usage/terms/dc/#UDC-002
Replaces:http://dublincore.org/usage/terms/dc/#UDC-001
```

The possibility to declare fully an external vocabulary is very important and with XML/RDF syntax DCMES promises to support fully any kind of external authority control. The issues and solutions of DCMES and UDC are linked to specific practical implementations and published reports are still very rare (cf. Slavic & Baiget, 2001).

[47] When classification is used it is advisable to provide natural language search terms as keywords. This approach is, for instance, used in DCMES in the EASEL project (see Appendix 3.1, Figure 3.1) (Slavic, 2003).

3.5.2 LOM - Learning Object Metadata

Maintenance and background. IEEE LOM is a metadata standard developed in 1997 by The Institute of Electrical and Electronics Engineers (IEEE) - Learning Technology Standards Committee (LTSC). This metadata standard was created to be used in education and training systems. It supports discovery, evaluation and management of learning objects, where a learning object is defined as any entity, digital or non-digital, which can be used, re-used or referenced during technology-supported learning.

Machine readability of metadata is paramount in the electronic learning environment and this has influenced the presentation of the LOM standard: data naming, formatting and detailed instruction for syntax and vocabularies. The standard itself was built with automated information management in mind. From the very beginning, XML schemas were made available to encode the standard (IMS Learning Resource Meta-data Information Model, 2001).

Structure. The elements of the LOM conceptual schema are organized into 9 functional groups that include not only descriptive metadata but also administrative, technical and education specific data. LOM can be used for different functions: to facilitate searching, evaluation, acquisition, and use of learning objects but also the sharing and exchange of learning material. Each one of the 9 groups of elements in LOM contain hierarchically structured elements and sub-elements.

Subject data. There is provision for a subject description in two groups *1 General* and *9 External Classification*. In group *1 General*, the subject can be described in the elements *keywords*, *description* and *coverage*. *Keywords* and *description* are supposed to be populated with uncontrolled, unstructured text.

Library classification can be efficiently used to describe the subject in group *9 External Classification*. In order to use this element to describe subject it is necessary to state 'discipline' in the element 'purpose' (see example below and example 3.2 in Appendix 3). Standard controlled vocabularies such as a classification are recommended to be used and the name of the standard is declared as a source. It is possible either to use several discipline taxonomies to describe a resource, or to choose to describe the resource's discipline through several taxonpaths within the same taxonomy, if the resource covers more than one discipline or sub-discipline.

In a LOM encoding schema, each level of the classification tree can be expressed within a metadata instance, however, this was originally meant for a very simple subject taxonomy. The most useful feature of the LOM approach is the provision for both a classification number and its description (caption).

```
<classification>
<purpose>discipline</purpose>
<taxonpath>
<source>UDC</source>
<taxon>
<id>6</id>
<entry>Applied sciences</entry>
</taxon>
<taxon>
<id>61</id>
<entry>Medicine</entry>
</taxon>
<taxon>
<id>611</id>
<entry>Anatomy</entry>
</taxon>
</taxonpath>
</classification>
```

(See also Appendix 3.2)

UDC was successfully used for subject browsing and searching in LOM metadata in the EASEL project repositories in 2000-2002 (Slavic, 2003). Since 2002, the educational metadata community has developed a machine-readable format VDEX, for vocabulary exchange that can be used for management of subject classification schemes and linked to LOM metadata (Appendix 1.3).

3.5.3 AMICO Data Dictionary- Art Museum Image Consortium

Background and maintenance. Museums have a tradition of maintaining different kinds of accession records (registers) purely for management and identification purposes and this kind of metadata is not used for resource discovery, collection browsing or retrieval. With automation in the 80s some bibliographic tools were introduced in the museum sector to provide a wider range of access points to the collection but this is not very widely accepted (Taylor A. G., 1999).

With the widespread use of the Internet, however, museum collections have become more exposed. The idea of virtual museums and gateways to information that include access to library, archive and museum collections, emphasises the rôle of museum data in resource discovery. Supporting resource discovery in the museum requires a different approach and differently structured metadata, as museum items are unique specimens by definition and they are viewed and organized in relation to their provenance.

Museum curators responsible for a collection, for instance, need to record richer and more structured information on the museum objects and this is obviously the direction for development of museum metadata standards which will accommodate both functions as was attempted with, for instance, the CIDOC[48] Reference Model (Doerr, 2003). The AMICO Data Dictionary schema was created to support museum object management issues by the Art Museum Image Consortium (http://www.amico.org) in 1999.[49] The schema is meant to be used with XML encoding.

Structure. The AMICO Data Dictionary consists of: Catalog Record Fields and Media Metadata Fields. Each work recorded in the AMICO Library is documented by a Catalog Record and this consists of seven groups of elements[50]:

AIDAmico identifier (Unique identification)
OTYObject type (What is it?)
OTGObject title/name (What is it called?)
OPDPhysical description (What does it look like?)
CTTCreator/Display/Text (Who made it?)
SUGSubject matter/Pre-iconographic description
OCRCritical responses (What does it mean?)

These main clusters are further structured into semantically related groups that contain a final number of descriptive attributes. Apart from element names, guidelines for use and suggested controlled vocabularies, each descriptive element is given an example.

Subject data. With careful consideration of what the subject data may be with regards to a museum object, it is obvious that there is no straight line between the subject of a written/verbal document and a museum object. Many facets of classification scheme vocabularies may, however, be applied in various AMICO metadata elements e.g. terms related to forms, size, materials, persons, space, time, events, art styles, periods, techniques, etc. Theoretically, these may be applied across the metadata schema e.g. within the **OPD- Physical description** in the Measurement

---

[48] CIDOC stands for International Committee for Documentation
[49] Other metadata standards for museums are *Computer Interchange of Museum Information* - CIMI Access Points, VRA Core (Visual Resources Association) and *Standard ProcEdures for CollecTions Recording Used in Museums* (SPECTRUM).
[50] These data elements are defined according to the Art Information Tasks Force / Getty core Categories for the Description of Works of Art (CDWA).

group, *Classification group*, *Object material group*, within the **CTT-Creator/Display/Text** in the *Style period group* and within the **SUG -Subject matter/Pre-iconographic description** in the *Context group* (Appendix 3.3).

So far the only standard indexing language suggested for use in the above mentioned element is the Art and Architecture Thesaurus. With respect to the differences in indexing museum objects as opposed to document like objects, it is evident that a classification such as UDC could equally serve the purpose. No application of UDC was reported in connection with AMICO Data Dictionary or indeed with any other museum specific metadata schemas.

3.5.4 EAD - Encoded Archival Description

Background and maintenance body. Archival material is arranged and described as a group of individual items with the same provenance and according to their original creation order. Archival metadata represents this hierarchical collection structure and organization. An item's description constitutes part of this structure.

Traditionally there are several categories of data: *donor file*, *accession data* (information about the source, donors, circumstances of acquisition, physical characteristics of the acquisition); and *finding aid* - a detailed description of the content of the archival records and their organizational context (Taylor A. G., 1999; Caplan, 2003). Since its official release in 1997, the EAD has become the metadata schema for encoding archival finding aids and is maintained in the Network Development and MARC Standards Office of the Library of Congress in partnership with the Society of American Archivists.

Structure. The EAD structure corresponds with archival practice in which attention is paid to the individual items in hierarchical arrangements, which allows for 'nesting' of a document within a document. A metadata schema consists of two basic groups of elements: 'meta-metadata' *elements* and *archival description* elements. Meta-metadata (or administrative metadata) provides information about finding the aid itself. *Archival Description* <archdesc> provides information about a body of archival materials. When created, it was envisaged that EAD would be encoded with robust and reliable Standard Generalized Markup Language (SGML). After the corresponding XML DTD was made available, this became a prevalent binding format for EAD applications.

Subject data. Subject description in EAD may be said to be reasonably well supported. It can be expressed in two ways, either as a <*subject*> element, or, when the subject is a person, corporate or geographic entity, it may be represented by elements <*persname*>, <*corpname*> and <*geoname*>. A <*subject*> element can be nested within 15 structural elements of the archival description. The subject scheme from which a controlled term is taken is declared as a source. The advantage is that this holds the possibility of linking to authority file data. The disadvantage is that some of the tags can be mixed with a free text subject description, which complicates their usability (Burnard & Light, 1996).

EAD metadata may exploit existing elements for a subject description in the <header> and <archdesc> group of elements. In the archival description data <archdesc>, the *Subject* element may be repeated for every item description for which this kind of indexing may prove useful. It can be used within elements such as Paragraph <p> for instance. If it is necessary to indicate a subject for the major representation of the material, the subject element can be nested within <archdesc><controlaccess>. The total list of elements that can nest *subject* is given as follows: *controlaccess, entry, event, extref, extrefloc, indexentry, item, label, namegrp, p, physdesc, physfacet, ref, refloc, unittitle.* (Appendix 3.4)

Due to its link to MARC formats, the EAD subject element is designed to be comparable to MARC fields for subject data 6xx (650 topical subject heading). EAD guidelines recommend the use of controlled vocabulary to facilitate access to subjects within and across finding aid systems. The attribute *source* is suggested as an addition to the element to specify vocabulary from which the term has been taken. If the EAD subject uses a term from an authority file, the attribute AUTHFILENUMBER can hold the link or the authority file record (Encoded Archival Description, 1998).

Access points relevant for the use of archival records are of a more formal nature and in favour of a collection approach. 'Aboutness' and 'topicality', however, become very important when an archival collection starts to be accessed via a larger portal or gateway service and this may influence future implementations. Classification as a controlled vocabulary is recommended, including UDC. The fact that the Library of Congress is the maintenance body and that EAD was initially applied in the USA makes EAD biased towards the choice of controlled vocabularies. So far, examples can be found for the use of LCSH and DDC. Although there is no evidence about the application of UDC in EAD there is no obstacle for this to happen in the future with the growing use of the EAD in Europe.

3.5.5 Summary on metadata and classification

Complex domain specific metadata schemas examined in this research reflect the policy towards subject description as established in information exchange in libraries, archives, museums and education. The fact that these domains interact over the Internet led to the creation of crosswalks and mappings to simple structured generic and domain independent schema such as DCMES and the need for sharing subject vocabularies (Baker, 2001; Doerr, 2003). All schemas described here suggest, in their guidelines, the use of library classification and explicitly recommend, for instance, the UDC (see Table 3.6 at the beginning of this section). The application of UDC outside the library domain is, however, still sparse and as discovered during the course of this research, only DCMES and LOM are actually being used with UDC in practical applications (Table 3.7). The analysis of a metadata standards shows that a UDC vocabulary may be potentially used as the vocabulary for many descriptive elements (type, form, format, audience) but none of them is known to have exploited UDC for purposes other than the subject.

Table 3.7: UDC in metadata formats

	DCMES	LOM	AMICO	EAD	MARC
ELEMENTS SUITABLE FOR CLASSIFICATION[51]	2 dc:subject dc: coverage	2 general classification (discipline)	5 groups of elements: Classification Group Object Material Group Style Period Group Subject matter Context group	15 <subject> can be nested in *controlaccess, entry, event, extref, extrefloc, indexentry, item, label, namegrp, p, physdesc, physfacet, ref, refloc, unittitle*	1 6xx group of elements for classification and 7xx for other systems
REPEATABLE	YES	YES	YES	YES	YES
IDENTIFIER OF SCHEME URI/TEXT	YES	YES	YES	YES	YES (UDC specific field)
IDENTIFIER OF A CONCEPT	YES URI	YES URI	YES URI	YES: "authfilenumber"	YES
LINK TO EXTERNAL VOCABULARY	YES, RECOMMENDED AND SYNTAX PROVIDED	YES	NO	YES	YES
SYNTAX SIMPLE/ STRUCTRED	SIMPLE	STRUCTURED	SIMPLE	SIMPLE	SIMPLE
USE OF UDC	YES	YES	NO	NO	YES

[51] All subject and other metadata elements where, for instance, UDC may be used.

The analysis of subject specific fields of the schemas shows that these are equally well or even better equipped to hold and to reference classification data than a library format. Research into the background and the domain specific policy towards metadata show that the choice of subject vocabulary is rooted in a domain tradition and only newly developed schemas such as DCMES and LOM are therefore likely to be open to experiment with UDC.

In relation to subject elements Table 3.7 shows that classification could potentially be used for subject description in more than one metadata element. EAD, being a hierarchically nested collection description, is exceptionally complex in that respect.

All five metadata standards allow for a subject vocabulary scheme to be optionally identified either through its name (free text) or through an URI. This means that any value used in the subject field ought to be interpreted against the scheme identifier. All metadata schemas allow for a single subject term to be both expressed and uniquely identified, which is paramount for the use of classification as an external vocabulary. The subject field is repeatable which means that more than one classification number and more than one subject vocabulary can be used in a single metadata instance.

From the point of view of element structure, only the LOM schema allows for hierarchical structuring of the classification number and makes provision for a classification caption and classification number to be placed together within the same element. The general impression is, however, that metadata schemas rely on external vocabulary to provide a structured interpretation of a given classmark and the searching or browsing function is obviously not expected to be based on a metadata instance but rather through a more complex metadata architecture.

The linking of metadata to an external vocabulary means that metadata elements holding the subject can remain simple, irrespective of the fact that, for instance, the classification number may be structured and may contain more information relevant for information retrieval. Figure 3.3 illustrates that in the same way that resources are kept separate from metadata, vocabulary used in metadata can be held separate from the metadata and this architecture allows for independent management and access control to all three parts of a networked information system.

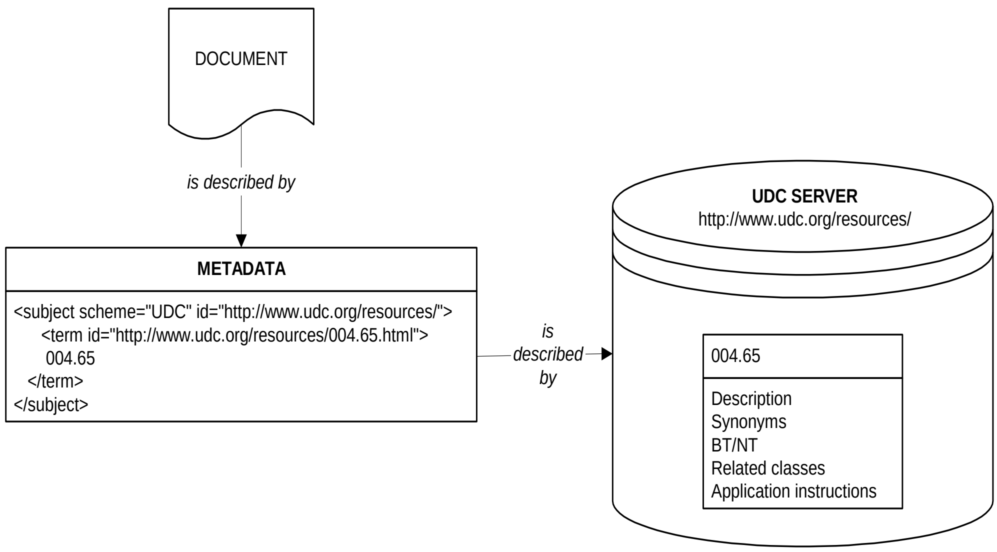


Figure 3.3: Metadata searching via external vocabularies

It is important to acknowledge the difference between using classification and any other alphabetical indexing language in metadata. Users on a remote search interface cannot understand classification notation if there is no link with a classification scheme or subject alphabetical index. An information service architecture, or more specifically, a metadata architecture will therefore play an important rôle in the use of classification.

In a networked environment the relationships between a classification authority file and the metadata subject element can be managed in different ways:

- use of URI/ID for external vocabulary, instead of UDC notation
- use both authority URI/ID and UDC notation
- use UDC notation locally but replace it with: notation, caption and URI/ID when exporting/exchanging metadata
- use UDC notation only, but export/exchange classification schedules together with metadata
- have only the UDC notation value in the metadata and use the authority file only as a cataloguing tool, source

Apart from the general feeling that the last solution, i.e. the use of UDC notation only, is not acceptable, there is no single recommendation or solution and the preference seems to be in having solution ii) in which both URI/ID and UDC notation is used (Powell, 2003).

One of the main concerns in metadata architecture, and in the Semantic Web infrastructure in general, is how to establish permanent identifiers for different vocabulary schemes so that they can be referred to from within metadata. Developments in vocabulary mark-up standards and Web ontology languages as described in Chapter One are in line with the Web architecture in which machine understandable thesauri and classification schemes would be made available and would be shared via the Web. For this to work every concept in every controlled vocabulary in this context needs to have a unique identifier that would enable its semantic interpretation, re-use and sharing between metadata schemas (Van de Sompel et al., 2004; cf. Vocabulary ML, 2000; see also NKOS web site at http://nkos.slis.kent.edu/).

The proposals for a vocabulary 'registry' and 'terminological services' are being discussed (Pepper & Garshol, 2002; Vizine-Goetz, 2003; Vizine-Goetz et al., 2004). One of the developments in the area of identifying vocabulary scheme concepts is the suggestion of a *Published Subject Identifier* (PSI):

> "*Published Subjects is an open, distributed mechanism for defining unique global identifiers. Based on URIs, the Published Subjects mechanism has two unique characteristics: it works from the bottom up, and it works for humans AND computers*" (Vatant, 2004: discussion list message; cf. Published subjects: introduction and basic requirements, 2003).

The use of a classification vocabulary in metadata, and in general, supporting resource discovery on the Internet will be closely tied to the form and format in which a given vocabulary is made available. As pointed out by Vizine-Goetz et al. (2004a) in order to create terminological services the existing vocabulary formats (e.g. MARC authority formats) will need to be enhanced with permanent identifiers and more data on term provenance.

3.6 Summary of UDC use on the Internet

The use of classification on the Internet, and the UDC specifically, started with uncoordinated efforts led by research projects, moving towards the organization of information on the Internet. At this stage, classification was used in its basic rôle as a browsing taxonomy (1-3 hierarchical levels and a few hundred classmarks). Medium and large size quality subject gateways were created to support resource discovery for the academic and research community. This was accompanied by the development of the necessary technological infrastructure to support information management and organization, i.e. metadata and other technical standards, protocols and tools.

Quality subject gateways have gradually become complex subject-orientated hubs offering cross-domain and cross-subject resource discovery, accompanied by current awareness and selective dissemination of information services. The primary goal here is cost effective and rapid resource indexing and delivery, using machine-assisted or fully automated tools. In addition, hubs and portals are built on different distributed collections that use different indexing systems and the value of classification in this environment is based on its capability of being mapped to other systems. The use of classification in current information services can no longer be judged by its 'naive' presentation on the interface but rather according to the rôle it plays in controlling term assignment and in providing a semantic framework for browsing and search expansion.

UDC was quite a popular classification system in the first phase of subject gateway development. After 1997 UDC seems to have become less attractive for research projects that give preference to vocabularies that are free for research purposes and for which the licence agreement is quick and easy to obtain and can be planned well in advance. The last project that did an extensive and advanced experiment with UDC was GERHARD. After the year 2000 UDC seems to be more popular in subject gateways in Central and Eastern Europe.

Because of its use in resource discovery, UDC was recommended as a standard vocabulary in many metadata standards. The use of classification and UDC in metadata is, however, still in an early phase of development. The important development in metadata architecture is its orientation towards an information architecture based on an external authority control. This kind of classification use requires standardized and machine-readable classification data and suggests the enhancement of the classification vocabulary that can support its better exploitation in the networked environment.

CHAPTER FOUR: UDC IN INFORMATION RETRIEVAL

4.1 Introduction

The reasons for using classifications in IR and the way classifications work in improving recall, precision and removing search disambiguation are illustrated in this chapter. The main characteristics of the use of UDC in IR are also explained in detail. Different approaches to using UDC for browsing in an online environment, along with the associated issues, are examined.

The difference between searching for words and searching for notation is explored. Access to UDC via words is the most important condition for using classification in a current online environment and different solutions are described. In order for complex notational expression to be searched in a post-coordinate manner (either with notation or words) complex notational syntax ought to be amenable to machine processing. Searching complex UDC numbers can be achieved using separate programs built on algorithms for splitting previously constructed complex notation. Alternatively, notation can be coded in the process of notation building in a way that will enable its processing. These two approaches will be explained.

4.1.1 Features of analytico-synthetic classifications

The analytico-synthetic nature of the UDC is based on the relational organization of vocabulary into separate vocabulary facets (where 'facet' is understood in the meaning of mutually exclusive hierarchies of concepts). This structure allows for re-use of concepts in different combinations to express detailed and complex subjects.

A pre-combined classmark is created in the process of indexing and this 'notation building' is facilitated by notational devices and set of rules to determine the combination of notational elements and their order (the citation order). In the majority of cases, pre-combined UDC classmarks preserve the original format of their data elements and they can be recognized and retrieved in a post-coordinated manner (see Appendix 10.2).

94"17"(4)	**History--18th century--Europe**
94"17"(430)	History--18th century--Germany
94"17"(436)	History--18th century--Austria
94"17"(44)	History--18th century--France
94"18"(4)	**History--19th century--Europe**
94"18"(430)	History--19th century--Germany
94"18"(436)	History--19th century--Austria
94"18"(44)	History--19th century--France
94"19"(4)	**History--20th century--Europe**
94"19"(430)	History--20th century--Germany
94"19"(436)	History--20th century--Austria
94"19"(44)	History--20th century--France

In the example above, the pre-coordinated classmarks allow for grouping and display of history by time and by country but also allow a post-coordinated search (by number or by words if these are made available):

94 AND (440)
[history] AND [France]

The post-coordinated search of pre-coordinated terms will retrieve an intersection of subjects but will also enable the logical display of results (in this case according to the period).

94"17"(44)	History of 18th century in France
94"18"(44)	History of 19th century in France
94"1800/1900"(44)	History from 1800 to 1900 in France
94"1804/1894"(44)	History from 1804 to1894 in France
94"1876/1899"(44)	History from 1876 to1899 in France
94"19"(44)	History of 20th century in France

Apart from special auxiliaries, all UDC numbers have exactly the same meaning regardless of their position in a composite UDC number. The ordering of numbers itself as regulated by different indexing practices, however, may add another layer of meaning to each UDC class, for instance, a 'treated subject' is often cited first followed by the 'subject of treatment', for example *33:622 The [influence] of mining on economics* as opposed to *622:33 The [influence] of economics on mining*. The basic meaning of a class as defined in the UDC schedules, however, is constant.[52] In the earlier example of history, one can choose the collocation of subjects pertinent for a certain collection , for example, by stating a *place* followed by a *time* e.g. *94(44)"18"*, and not *94"18"(44)*, in which case the initial collocation would be by country and not by century.

This 'subject building' where elements of compound and complex subjects are represented with discrete data elements is especially useful and easy to implement when UDC is used in the field of humanities. Here information organization and seeking is often related not only to a topical subject combined with temporal and spatial coverage but also to a document's external form and language. This is very much the case for history (as shown above), philosophy (e.g. Indian, Greek, German Philosophy, in a certain time, and the document in the original language or a translation), arts, literature and religion.

Using UDC, for instance, it is easy to make distinctions between English literature of Africa, America, Australia, Canada, India, by time or by genre, criticism or original work. Alternatively it is possible to collocate by search e.g. African or Asian literature written in e.g. English, French, German etc., as opposed to African literature written in indigenous languages. Furthermore, it is possible to search, for instance, African indigenous poetry available in English or French translations. This particular type of detailed document indexing that allows post-coordinate searching is not easy to achieve with enumerative classifications.

The most important feature of the UDC, however, is that it allows for the combination of two distinct subjects that form the content of a single document using a colon, such as *37:32* - the relationship between *education* and *politics*. Post-coordinated searching of these two elements *37* AND *32* gives the same precision in searching as any other post-coordinated indexing language. This search will retrieve the intersection of these two numbers in any of the following combinations:

---

[52] Typically for any pre-coordinated system a different order of concepts in combination can be used to denote different subjects e.g. *75"18"(44) Painting in 19th century in France* (may include Italian, Dutch etc. painters and painting that exhibited/worked in France in 19th century), as opposed to *75(44)"18" French painting in 19th century*. When these numbers are searched in a post-coordinate fashion this difference in meaning is lost.

32+37	Education and Politics
32/37	range of subjects in the social sciences from Politics to Education
32:37	[relation between] Politics and Education
32:37:2	[relation between] Politics, Education and Religion

This is why pre-coordinated languages are often said to have the ability to enhance precision, providing specificity, suggestibility and contextuality without sacrificing recall (Svenonius, 1993; Bodoff & Kambil, 1998).

4.1.2 Contextualisation and disambiguation

If UDC is fully implemented it can provide a good level of support for searching using notations or words. Concepts are semantically and syntactically organized and linked within a classification structure, which is represented using decimal notation. Each class notation is described using a caption (text) and additional search terms can be associated with each class through a subject-alphabetical index.

When detached from the classification structure, a word may be ambiguous and users may not be aware that the terms they search for may have other meanings. In any such case, classification can help resolve homonyms and polysemes by showing their position in the knowledge structure. For instance, the terms *cell* and *structure* appear together in UDC in two different subject contexts:

576.3	General cytology. The cell as a biological system [Biology]
69.056.56	Cell structures. Box structures (room-size cells) [Building]

Vickery (1990) stresses that a hierarchical classification structure can support subject queries to guide subject indication, term selection, to infer subjects, to link query terms to index terms and to disambiguate terms. More specifically, the combination of natural language terms and classification structure can be exploited by an IR system to:

- provide an interactive two step search allowing users to choose the right subject area by offering classified results
- search a combination of terms against a certain class
- allow users to choose search terms from UDC based dictionary

The problem of disambiguation within a search can be resolved by displaying a class description with the class number and also by enriching the classification file with search terms and synonyms. This function is directly linked to the quality and explanatory power of the UDC number description (caption) and the availability of a subject-alphabetical index to the classification.

Synonyms may create a problem in word searching.[53] Should potential search terms and variants of terms not appear anywhere in a caption, they need to be added to the system in sufficient numbers, pointing to the right place in the schedules. Unlike past printed editions, in a database there is no limitation to the number of synonyms one can add as an access point to a class.[54] When it comes to homonyms and polysemes i.e. identical terms with different/unrelated and different/related meanings these can occur in a caption in different places in the schedules. In this case a subject-alphabetical index collocates different subjects under terms that may look the same, prompting the user to choose between them.

4.1.3 Improving precision

Precision in retrieval can be achieved straightforwardly through logical coordination of UDC numbers in the searching process (Boolean logical operators):

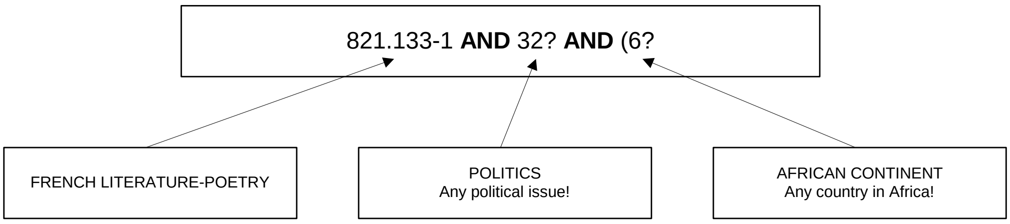


Figure 4.1: Post-coordinate UDC searching

The search expression in Figure 4.1 looks for documents on the influence of politics on *Fracophone African poetry*. This search will exclude any other literary genre and also the poetry of African authors written in an indigenous or any other language and will thus contribute to the search precision. The truncation of *32?*

[53] Classification is known to be context and language independent as it implicitly solves the problems of synonyms, but this is only if one searches for notations. Any application of classification in IR today assumes that users will not be using notations but rather words.
[54] This is a weakness of the UDC MRF which is distributed to users without a subject-alphabetical index and the schedules themselves often do not contain enough index terms to support their straightforward use in IR.

Politics, on the other hand, will allow for any political issue influencing literature to be included and the truncation of *(6? Africa* will ensure that any African country meets the criterion. Alternatively, a political issue or an African country can be specified if necessary.

Precision in retrieval using UDC will depend upon the precision and depth of indexing. This is usually achieved by choosing the most specific subject in one of the following ways:

- using the most specific classes in the hierarchy e.g., choosing *621.311.212 Run-of-river-power stations* instead of *621.311.21 Hydroelectric power stations* or the even broader *621.311.2 Power stations according to type of prime mover*.
- using a combination of numbers i.e. expressing the subject using compound (or complex) numbers:

  e.g. *621.311.212.01 General studies, theory of run-of-power* (combination of main number with a special auxiliary)

  621.311.212"18"(410)(047.31) Run-of-river power stations in 19th century in England, research report (combination of main number and common auxiliaries of time, place, form)
- using a combination of two main numbers with a relation symbol *621.311.212:502 The [influence] of environmental science on the run-of-river power stations* (combining two main numbers with a relation symbol)

A UDC notation can express syntactical, i.e. grammatical relationships using symbols: a relation is denoted by the colon **:** , double colon **::** , addition by a plus + , subordination with square brackets [ ] and a range of successive subjects with an oblique stroke (forward slash) /. The possibility of expressing a combination of different subjects within a single number is a powerful device in indexing.

The colon combination of two UDC numbers may, however, denote different relationships e.g. 'influence', 'application' or 'comparison' between subjects. At the same time, if notations in a colon combination follow the order in which they occur in the table (as suggested by McIlwaine, 2000b: 37) it is impossible to determine which is the treated subject and which is the subject of treatment (for example does

32:37 mean *the influence/application of politics on education* or *the influence of education on politics*).

Further precision could be easily achieved by adding distinct symbols for different relationships and by adopting the rule for ordering subjects within relationships. Research done by Perreault (1969, 1969a, 1994) shows that UDC syntax could be further improved to enhance precision in indexing. He suggested that a combination of UDC numbers, expressed by a colon symbol, should be refined and fully expressive of different logical relationships.

The comprehensive list of relationships between concepts given by Perreault (1978) can be outlined in six main categories: ordinality (preceding to... less than, smaller than); determination (provoking..., giving birth to, limiting...); attribution (characterised by...); interactivity (different from..., concordant with..., imitating...); subsumption (having parts such as..., being of kind such as...); logic (negation of...., reciprocity to .., opposite from) (Perreault, 1969).

In 2003 it was, however, decided that phase relationships between subjects should be expressed through the use of common auxiliary numbers *-04 Relations, processes and operations* (and not by the refinement of existing symbols). The new common auxiliary table offers 11 common auxiliary numbers for phase relationships: bias phase, comparison phase, influence phase, tool phase (exposition phase), isolation, separation, association, independence, dependence, interdependence and complementarity.

Thus, although no explanation or example on how this is meant to be used are provided in UDC MRF 2003 and 2004, it appears that these common auxiliary numbers are meant to be used to denote a phase relationship between two UDC concepts by 'intercalating' a common auxiliary number between them e.g. instead of *37:17 [relationships between] Education and Ethics* it may be possible to state *37-**042.3**:17 Influence of Ethics on Religion*. The use of symbols for one kind of relationship (i.e. +, /, :, :: []), while using common auxiliary numbers for another (i.e. -04), however, does not seem to contribute to the structural and operational coherence of the UDC.

4.1.4 Improving recall

The classification potential in improving recall lies in the semantics, i.e. its hierarchical structure, and the fact that it groups subjects into relatively broad classes and subclasses (Svenonius, 1983). Dreese (1979), for instance, reported on a pilot study on retrieval efficiency comparing UDC with catchwords extracted from a document title that clearly demonstrated the strengths of classification in improving both recall and precision. In the majority of cases expansion of a search can be achieved through simple truncation of the search expression or by moving up the hierarchical tree on the browsing interface, and also by excluding parts of the compound number:

```
Search exact match:
821.111.09"18"=133.1
[English literature -- criticism studies -- 19th century in French]

broadening the search by eliminating a part of the combination:

821.111?-1"18"
821.111.09?"18"
```

```
Search exact match:
004.65
will give 004.65 Database management systems (DBMS)

broadening search with right truncation >>004.65? will give:

004.651   Database file organization
004.651.2        Partitioned organization
004.651.3        Multilist organization
004.651.4        Tree structure organization
004.651.5        Dynamic file organization
004.651.52              B-tree
004.651.53              Self-organization tree
004.651.54              Dynamic hash organization
004.652   Database models
004.652.2        Hierarchical model
004.652.3        Network model
004.652.4        Relational model
004.652.42              Relational integrity
004.652.43              Relational algebra
004.652.44              Relational calculus
004.652.5        Object-oriented data model
004.652.6        Logical database model
```

Improving recall by right truncation has often been mentioned in the literature as one of the main advantages when using decimal classifications in IR (Svenonius, 1983; Cochrane & Markey, 1985). In 1987 Markey Drabenstott (1987) reported an IR experiment using a specifically designed system: DOC - Dewey Online Catalog (in four participating libraries) to evaluate the use of classification in information retrieval. DOC was adjusted to match user queries with a broader class, although more specific classes matched the topic. The reason for this was the library practice of assigning broad class numbers even when DDC could offer a more specific subject. The DOC system allowed users to browse the broader class in case they obtained too few results on a specific search and in many cases this enabled users to find more desirable items.

The next IR experiment undertaken by Markey Drabenstott (1996) on a large bibliographic database using LCC and DDC shows that there is room for caution in making hasty generalisations. Namely, while the experiment results indicate very good performance of the two classifications in improving precision (i.e. reducing high posted searches), the expansion of searches by truncation to improve recall failed to produce the expected results, because the notation, on its own, does not always express hierarchy.

In relation to right truncation, used to broaden a search, the literature on UDC also reports a problem when right truncation does not lead to a broader class (Buxton, 1990; Riesthuis, 1998). This problem normally occurs in decimal notation whenever a subdivision of a broader class is centesimal or spans several numbers on the same level of division in which case a broader class can be expressed only as a range number. For instance, truncating the number *594.151 Cockles, 594.15?* will not lead to the broader class *594.14/.17 Heterodonta*:

594.14/.17	Heterodonta
594.141	Nayadidae. Unionidae. Freshwater mussels
594.151	Cardiidae. Cockles
594.18/.19	Desemodonta
594.185	Myidae. Gapers (soft clams)
594.188	Pholadidae. Teredo. Ship-worm

This is one of the reasons why search expansion should not be based on the UDC notation only. As pointed out by Loth (1996), in order to support search expansion, classification database formats require additional data on the correct broader class for each notation, so that a class hierarchy can be automatically and correctly managed independently of the notation itself. This will be explained later in Chapter Six.

4.2 UDC specific IR functions: advantages and issues

As UDC is both an indexing language using symbols, and a pre-coordinated indexing system, it has certain requirements in its management and application. Firstly, the link between UDC numbers and natural language terms has to be supported and properly controlled. Secondly, the classification data have to support the management of hierarchical structures and syntactic sequences in pre-combined UDC numbers.

4.2.1 UDC browsing

In indexing, classification aggregates documents by the degree of their similarity (logical grouping) whilst enabling semantic relationships between concepts and subjects to be presented in a visual manner. The class number given to a document goes no further than asserting that a document belongs to a class of documents that deal with the same class of subjects. Or, as Perreault (1990: 18) explains, a document is prescribed to one class among many, not described by it.

Fugmann (1990: 67) emphasises the importance of category-based indexing as exemplified by classifications and suggests that they "guarantee just that degree of representational predictability and fidelity which a language displays and which one expects of it as an information system user". Hence, classification will aggregate documents dealing with the same subject field at different levels of specificity while alphabetical indexing will result in a scattering of the subject field across the alphabetical range of terms. Classifications are good for subject browsing, alphabetical indexing languages are more appropriate for searching, and therefore both are meant to complement one another in information retrieval. Classification supports browsing as it provides a hierarchically organized concept space that can be represented as a nested, indented, hierarchical subject tree. In addition, through *see also* references concepts distributed across different hierarchies are semantically associated. A classification's value is in visualising semantic relationships: hierarchical (subordinate, superordinate), associative (collateral) and coordinated (e.g. the range of coordinated hierarchical subdivisions).[55]

---

[55] Bhattacharyya suggests the abbreviation "COSSCO" to summarise all the relationships in a classification system: coordinated, superordinated, subordinated and collateral (Bhattacharyya, 1979).

Functional requirements for the effective browsing of a system have been defined as: the option given to users to position themselves in an area of interest, the ability to recognize an appropriate direction in which to proceed (how to broaden and narrow), and the ability to move easily through the database (Cox, 1992: 4). Effectively, this means the easy transition from search to browse, where a search has an important function in providing entry points to browsing.

When talking about intelligent interfaces, Vickery (1990) suggests that every interface should offer a dictionary in the form of a classification or semantic network, or an alphabetical list supplemented by classification. Entry points to browsing are very important and should be suggested and made obvious by the system. Although one can start browsing by random selection, or by opportunistic examination (top/down browsing), or by a search from the subject index, it is mainly the last that is most important for large classification systems.

Needless to say, every classification-browsing interface should consist of a caption display with or without UDC notation. Logically the display of notation alongside captions in the case of UDC will help to visualize hierarchical relationships. When notations are removed from the users display, captions have to be indented to show the hierarchical relationships. In any case, it is not considered of good practice to display UDC numbers without their caption.[56]

Associative relationships in classification (*see also* references) may link different aspects of a subject and this can be implemented in information retrieval to support subject exploration and expansion through lateral browsing. A reference may need to be linked to several steps of association.

Figure 4.2 shows class *316.35* containing three references: *061*, *316.342.6*, and *316.45*, that, if activated as links, may lead to new subjects. The number *061* itself has two further references: *347.19* and *347.471*. In turn, *347.471* has reference *334.7*.

[56] Displaying classification numbers on the user interface without a description is a typical fault of library catalogues.

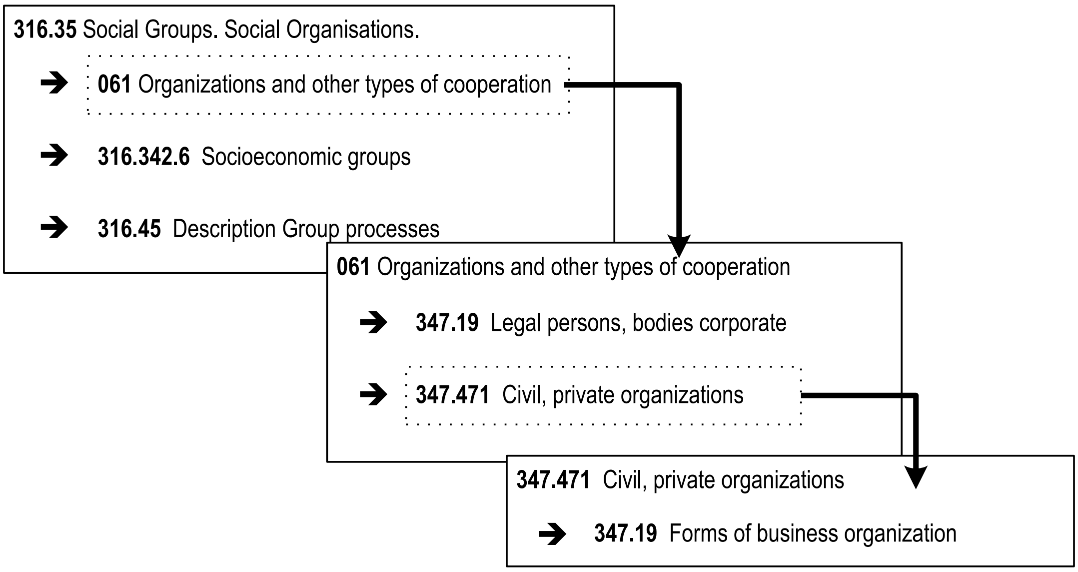


Figure 4.2: Lateral browsing through '*see also'* references

In the process of browsing, a fully expanded UDC classification hierarchy often may be too detailed, comprehensive and difficult for users to gain an overview of a subject area and the direction in which to browse. This is why it is important to be able to 'collapse' and 'expand' a hierarchy. By collapsing hierarchy, coordinate classes become clearer and help users to position themselves for further browsing or to broaden/narrow the search, as is shown in Figure 4.3.

332.1	**Regional economics. Territorial economics**
332.12	**Regional, territorial structure of the economy**
332.13	**Location, regional distribution of factors of production...**
332.14	**Regional, terrotorial economic policy, planning**
332.15	**Selection of location. Site selection for individual economic...**
332.2	**Land economics**

332.1	**Regional economics. Territorial economics**
332.12	**Regional, territorial structure of the economy**
332.122	Economic areas. Economic regions
332.122.5	Economic areas according to thei homogenuity
332.122.52	Homogeneous regions
332.122.54	Heterogeneous regions
332.122.6	Economic areas according to their function
332.122.62	Function zones
332.13	**Location, regional distribution of factors of production...**
332.14	**Regional, terrotorial economic policy, planning**
332.15	**Selection of location. Site selection for individual economic...**
332.2	**Land economics**

Figure 4.3: Collapsing hierarchy to gain overview of the subject field

Issues related to the interface to a classification are going to be addressed in more detail in Chapter Five.

4.2.1.1 Presenting UDC in a linear order

The UDC structure and syntax function in such a way that the classification always organizes subjects from general to specific. Subject representation is formalised and 'fixed' through decimal notation combined with symbols. The correct filing of notation will hold the knowledge structure in place and thus allow subject display for the purpose of browsing.

Each UDC notation contains an indicator of the table from which it originates (usually a combination of brackets, digits, symbols and punctuation or simply the absence of any of those) and this 'indicator' establishes a sequence for notation filing. The filing order of classes is thus defined by the meaning of the symbols rather than by the simple mechanical ordering of notation, which is why the UDC filing order is said to be intellectual and not mechanical (Figure 4.4).

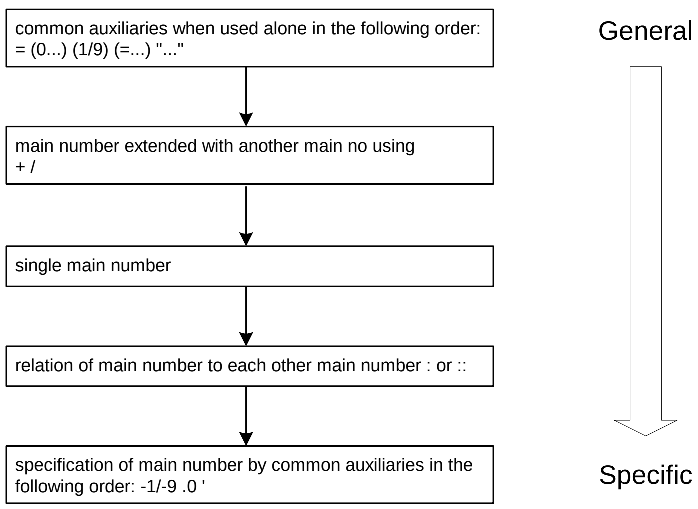


Figure 4.4: General to specific filing order

When applied, this produces the following filing sequence:

73+75	subject: Plastic Arts and Painting
73/75	subject: 73 Plastic Arts and 74 Drawing and 75 Painting
73	subject: Plastic Arts
73:74	subject: Plastic Arts in relation to Drawing
73=111	subject: Plastic Arts- document in English language
73(410)	subject: Plastic Arts in UK
73"18"(410)	subject: Plastic Arts in 19th century in UK

Single numbers follow the rule of decimal number ordering:

328	Parliaments. Representation of the people. Governments
328.1	Parliament and government
328.17	Resignation
328.172	Resignation of cabinet members
328.3	Function of parliament

Pre-combined UDC numbers in the example below follow the rule of ordering of combinations as shown earlier in Figure 4.4:

328	Parliaments. Representation of the people. Governments
328"17"	Parliaments in 18th century
328"18"	Parliaments in 19th century
328"18"(4)	Parliaments in 19th century in Europe
328.1	Parliament and government
328.1"18"(410)	Parliament and government in 19th century in England
328.1"18"(44)	Parliament and government in 19th century in France
328.1"18"(45)	Parliament and government in 19th century in Italy
328.1"18"(5)	Parliament and government in 19th century in Asia
328.1(410)(091)	History of parliament and government in England
328.1(44)(091)=40	History of Parliament and government in France. In French language
328.17	Resignation
328.172	Resignation of cabinet members
328.3	Function of parliament

In an online environment there is a discrepancy between desired, arbitrary UDC order and computer filing based on character coding tables such as American Standard Code for Information Interchange (ASCII) which results in the following discrepancy:

Table 4.1: ASCII filing of UDC notation

ASCII order	Desired order
398"18"	398=111
398(=111)	398(042)
398(042)	398(410)
398(410)	398(=111)
398*	398"18"
398.01	398*
398=111	398MEA
398'01	398-02
398-02	398-03
398-03	398-05
398-05	398-1
398-1	398.01
398MEA	398'01

This affects not only combinations of numbers with /, + : combinations, but direct combinations between main numbers and special auxiliaries and main numbers and common auxiliaries. The problem is not peculiar to UDC; other library classifications, created for manual handling are also affected.

The incorrect filing of call numbers using DDC and Cutter numbers (a combination of numbers and letters used to achieve an alphabetical sequence within a class) or the LCC (a notation consisting of numbers and letters) has been subject of frequent complaints (Cochrane & Markey, 1985; Chan, 1986; Howard, 1990; Kniesner & Willman, 1995). Howard (1990), for instance, suggested a change of character code table and the creation of transformational tables for library systems to help with DDC and LCC systems.

The change of the character code table may affect other parts of the system and still may not help in ordering numbers connected with /, : or + which is an issue particular to the UDC. In addition, special programs or a transformation table for sorting numbers works within a single system but has no benefits in information exchange. Instead of changing the computer filing sequence, the better approach would be to change UDC symbols to correspond a computer filing order. Freeman & Atherton (1969), for instance, replaced UDC notational symbols with letters for marking tables which resulted in a full and correct mechanisation of filing, as shown in Table 4.2 (based on Freeman, 1964: 140).

Table 4.2: Freeman's set of substitutes for UDC symbols

DISPLAY	SUBSTITUTE CHARACTERS	EXAMPLE	INSTEAD OF
X+Y	X**A**Y	73**A**75	73+75
X/Y	X**B**Y	73**B**75	73/75
X	X**C**	73**C**	73
X:Y	X**D**Y	73**D**74	73:75
X[Y]	X**E**Y	73**E**371.3D7	73[371.3:7
X=Y	X**E**Y	73**F**111	73=111
X"Y"	X**G**Y	73**G**18	73"18"
X(Y)	X**H**Y	73**H**410	73(410)
X(=Y)	X**J**Y	73**J**111	73(=111)
XABC	X**K**ABC	73**K**ROD	73ROD
X.00Y	X**L**Y	73**L**1	73.001
X-Y	X**M**Y	73**M**03	73-03
X.0Y	X**N**Y	73**N**032.5	73.032.5
X'Y	X**P**Y	not in use with 73	not in use with 73

Dahlberg (1971) later suggested that this solution be implemented in the UDC schedules. Although this suggestion would make UDC significantly easier to implement and use in automated systems this change was never widely adopted. The problem was later revisited by Buxton (1990) and he suggested this solution as practical and desirable when UDC was used online. The ordering of auxiliaries in computer systems using letters and substitutes for non-conventional symbols is a very simple way to solve the discrepancy between intellectual ordering and mechanical filing.

Another approach is to replace UDC symbols in the database either by alternative coding or sub-field coding, and to use the original UDC symbols for display only. Sorting would be achieved through codes and tags and not through the text of the UDC notation. This requires management using a classification authority file. A list of codes proposed for the UDC MRF field *v004 UDC combination type* in 1993 could be used for such a purpose, as shown in Table 4.3 (see Master Reference File Manual, 2003):

Table 4.3: Scheme for coding UDC notation based on MRF Manual

MRF COMBINATIONS TYPE	COMBINATION CODING BASED ON MRF
a Plus addition [+]	X**a**Y
b Stroke range [/]	X**b**Y
c Simple number	X**c**
d Intercalation	X**d**
e Colon [:] combination	X**e**Y
f Double colon [::] combination	X**f**Y
g Combination with Com. Aux. of Lang. [=...]	X**g**Y
h Combination Com. Aux. of Form [(0...)]	X**h**Y
i Combination with Com. Aux. of Place [(1/9)]	X**i**Y
k Combination with Com. Aux. of Ethnic Grouping [(=...)]	X**k**Y
l Combination with Com. Aux. of Time ["..."]	X**l**Y
m Combination with non-UDC notation [*]	X**m**Y
n Combination with A/Z	X**n**Y
o Combination with Com. Aux. of Point of View [.00]	X**o**Y
p Combination with -0 Com. Aux. Of Persons / Materials / Properties / Processes	X**p**Y
q Combination with hyphen (-) spec. aux.	X**q**Y
r Combination with point-nought (.0) spec. aux.	X**r**Y
s Combination with apostrophe (') spec. aux.	X**s**Y

In modern IR, controlled vocabularies are normally managed through authority files and there is no real reason for UDC to be filed improperly either by replacing symbols with letters or by managing the classification through the coding of table indicators and syntax elements so that the number filing is managed independently of its display. Issues related to the UDC authority file will be addressed in Chapter Six.

4.2.2 Searching UDC

Searching a classification includes searching the notations, words attached to the notations and a combination of words and notations. Searching UDC notations or a combination of words and notations is suitable for professionals, indexers and cataloguers familiar with the system. Searching a classification using words differs from searching via any other alphabetical indexing system since the classification behind words will allow:

- easy expansion of searching based on the classificatory structure (broadening or narrowing the search or expansion via class association)
- subject collocation and hierarchical display of results

The artificiality of a classification notation has both advantages and disadvantages in information retrieval. As the notation is formalised and coded, it may be suitable for computer systems but remain inadequate for use by end users. In a multilingual database, numeric notation is considered to be a useful advantage in performing the rôle of a 'universal' indexing language or in acting as an intermediate language between different retrieval languages.

In the history of UDC automation and especially with library OPACs, often the only way of making classification useful for end-users was to provide a facility for searching notations (Foskett, 1973; Marcella & Newton, 1995; Buxton, 1990). In modern IR systems it is both unnecessary and unacceptable to expose end-users to notation searching, and classification is expected to be linked and used via verbal expressions. There are situations, however, in which classification notation searching can be used for search expansion. This is still relevant for systems which use classification in a more primitive way, i.e. with no authority control for classification data. The characteristics and requirements for both approaches are explained in this section.

4.2.2.1 Searching notations

Searching notations, even when they are treated as simple text, can provide functionality that is hard to achieve by searching with any other alphabetical indexing language. For instance, simple truncation of notations has often been exemplified in the literature as the easiest way to broaden a search and a meaningful way to improve recall. Foskett (1973:80) illustrated a typical scenario with an example from the American Geological Institute of a document about a meteorite that fell near the Leicestershire village of *Barwell* in 1965. Using words, this document could be found only if the user knew the name of the village - looking for Leicestershire would not retrieve it. If the UDC number for Leicestershire is truncated: *(425.42?*, this would retrieve documents on all places in Leicestershire, including *Barwell*.

Searching UDC notations is a frequent practice among professionals in a European multilingual, multiscript environment, especially in Central and Eastern Europe (Loth & Funk, 1990; Riesthuis, 2003a). The advantage, described above, in search expansion based on classification structure can, however, be implemented irrespective of whether users search by words or by notations. It can be achieved through the management of UDC using a classification authority file in which both hierarchical relationships and links between notations and words can be easily controlled and maintained.

4.2.2.1.1 Decomposition of UDC numbers

When UDC numbers are entered in the system as simple text strings, as has often been the case in library catalogues, it is not possible to search for individual elements of pre-combined numbers. In this case, meaningful elements of pre-combined numbers are not detached or marked in any machine-readable way. For instance, UDC notation *37:32* [*The relationship between] Education and Politics* and *94"18"(410) History of UK in 19th century* would be given a bibliographic description as follows:

675	$a**37:32**

675	$a**94"18"(410)**

Therefore, searching for elements of a pre-combined notation such as *32* in the first example or *(410)* in the second example is not possible.

Based on the assumption that it would not be acceptable to cataloguers to enter classification data into a catalogue in a structured way, Liu (1996) who worked on DDC, suggested that the only way around this is to devise decomposition algorithms which would allow a search of pre-composed elements even if the classification numbers are held in the system as a simple text string. However, he reported on the difficulties in devising algorithms for 'decomposing' DDC numbers because of the lack of consistency between notation and classification structure.

Riesthuis did similar work in order to convert pre-coordinated into post-coordinated UDC notation.[57] But unlike Liu's experience, in the case of UDC, the process of deconstructing complex numbers to basic numbers based on facet indicators and syntax symbols was achievable. Riesthuis did his research on the data catalogue of *Biblioteca Centralã Universitară* in Bucharest which uses UDC for very detailed indexing and had access to around 175,000 notation combinations. He distinguished the following situations:

1. a notation from the main table or an independent common auxiliary table
2. a notation formed from a notation from the main table or an independent common auxiliary table by means of parallel subdivision
3. one of the first two cases, followed by one or more special auxiliaries; if the relevant common auxiliary table has an end symbol, a special auxiliary comes before the end symbol
4. a range of notations connected with /, with or without one or more special auxiliaries
5. one of the first four cases, combined with one or more notations from a general auxiliary table (which can have their own special auxiliaries)
6. two or more of the first five cases, combined with each other and :, + or ::.

(Riesthuis, 1999: 25)

To resolve these situations, Riesthuis defined seven groups of algorithms for deconstruction:

1. An algorithm that splits up subject notations with more than one notation from the main table, combined with :, :: or + into subject notations with

[57] Riesthuis' research was reported in his doctoral thesis (1998b) and number of research papers (Riesthuis, 1997, 1998, 1998a, 1999).

each having one notation from the main table

2. A group of algorithms which split up notations consisting of a main number and one or more general auxiliaries into the main number and the general auxiliaries. Special auxiliaries remain with the main numbers or the general auxiliaries to which they belong
3. A group of algorithms which split up combinations of the general auxiliaries with :, ::, and +
4. A group of algorithms which treat ranges of numbers with /
5. A group of algorithms for final digits
6. A group of algorithms for notations which are the result of parallel subdivision
7. A group of algorithms for notations with special auxiliaries

An additional group of algorithms is devised for searching text attached to a notation. The first three groups of algorithms can produce search elements without controlling the UDC schedules. The other algorithms, including the eighth, can work only if controlled against the UDC schedules (Riesthuis, 1998).

This analysis can serve as the basis for creating a computer application for searching elements of the UDC in a post-coordinate manner but also in relating these elements to their verbal explanation which was the main goal of Riesthuis' research. Namely, in order to automate searching UDC using words from the caption, it is necessary that all pre-combined numbers (that do not exist in the UDC MRF and, therefore, do not have a caption) be split into simple numbers that do exist in the database and contain a verbal description (Riesthuis, 1997).

Riesthuis's solution may be useful in systems that have limitations in the management of UDC data, i.e. do not have the possibility of holding structured UDC numbers in authority files. His research demonstrated the importance of a well formed and expressive notation and use of the strict UDC 'grammar' that can be expressed through algorithms. In order for his approach to work it is necessary to reduce the number of range notations or notation created through parallel division as these are not easy to control by decomposition algorithms.

The proposed decomposition of UDC notations for the purpose of IR, requires:

i) a decomposition program (such as the one created by Riesthuis)

ii) UDC MRF data in a format compatible with the decomposition program
iii) middleware which would connect these components with a library and/or IR system

None of the above mentioned requirements are available, compatible and ready to use, for instance, with library systems, neither are they free. But most importantly, every individual system using UDC has to invest the same effort and incur the same costs in order to be able to search notational text strings. Since there is no noteworthy IR system today that does not have the possibility for authority control and a facility for sharing the same classification data, the whole problem of searching UDC notation should be viewed in a different light as will be explained in Chapter Six.

4.2.2.2 Searching UDC using words

Searching a classification via verbal expressions was recognized as a basic requirement for the use of the UDC in IR as it allows users to search the UDC without prior knowledge of the classification. From the point of view of an IR system, the essential difference, as pointed out by Ścibor (1997), is whether: *(a)* verbal expressions linked to the UDC are used as an interface only and are automatically translated into numbers, hence the system operates with UDC numbers; or *(b)* whether documents are indexed simultaneously by UDC and by verbal expressions and retrieval is performed on words only. In the first case, the classification structure is preserved or, as Ścibor suggests, this can be transformed into a thesaurus built upon the UDC. In the second scenario, the classification hierarchical structure ceases to be a part of an IR system and, since it is not exploited in browsing or search expansion, it is not of interest here.

Searching UDC via words depends on separate management of classification data and a database tool that will maintain links between the classification notation and verbal expressions. In the online environment the relationship between words and symbols is anticipated to be managed automatically in the process of database interrogation.

When UDC schedules were made available as a database format in 1993, the UDC data became easier to implement for IR. It was chosen, however, to maintain and distribute UDC MRF without a subject-alphabetical index. As a consequence, the research and development into UDC searching using verbal expressions in 1993-

2005 falls into the following categories:

i) using UDC captions to provide access points via words
ii) creation of online subject-alphabetical index (simple or in the form of a thesaurus, monolingual or multilingual)
iii) mapping UDC to alphabetical indexing languages

When UDC schedules are used as the only source for verbal expression it is advisable to make not only UDC captions searchable but also other textual fields that may be a source of verbal expressions: the text of scope and application notes, and especially the descriptions of examples of combination. Success in IR in this case entirely depends on the availability of useful terms in the schedules. The non-specific caption, i.e. those that can be understood only through access to the class above, will lead to information loss (e.g. captions such as *Other parts*, *Others*, *General*).[58] The weakness of caption searching was realised in the early stages of automation and has influenced the revision process. UDC has been gradually reducing the number of class descriptions that have no meaning, which means that the UDC MRF text can be used to a reasonable level in both a direct search or as a source of terms in creating proper subject-alphabetical indexes.

UDC schedules (caption and notes), however, still contain a number of class descriptions that are understandable only in the context of a printed page. Research on automatic term extraction from UDC schedules highlights these and other problems that need to be dealt with when creating a search terminology (Riesthuis & Bliedung, 1990, 1991; Riesthuis, 1999; Kumar & Parameswaran, 1999).

The UDC's value in a multilingual, networked environment can be exploited only when translation of the classification into different languages is made functional in resource discovery. The solution may consist of an extraction of terms from the schedules (from editions in one or several languages) and making them available as access points to classification on the user interface. The other approach is mapping the UDC to one of the existing subject alphabetical languages in the language supported by the end-user interface. These two approaches are not mutually

[58] The same problem was reported for LCC and DDC classification when they are searched using caption (Cochrane & Johnson, 1996).

exclusive and can be merged at any point when the UDC is applied in a networked environment. For instance, UDC can start as the source for an automatic extraction of terms in one language. It can then be expanded with terms in several languages, and then it can become a base for the creation of a subject heading system or thesaurus and subsequently mapped to other classification systems or subject heading systems or thesauri.

Implementers of the UDC are most concerned about the costs of the whole undertaking and the functionality they get with different approaches. The cheapest way is to use the text of the UDC MRF to create an index term file and link it to the collection database that is to be searched. The procedure could to a large extent be automated. A more expensive way would be to create a proper subject-alphabetical index to the collection, which can again acquire different forms and have a different level of vocabulary control. The cost incurred by creating an index is the reason why in choosing a classification, the advantage will always be given to systems which contain professionally made subject-alphabetical indexes.

The value of a classification subject-alphabetical index is that it collocates distributed relatives i.e. related subjects that are scattered in the schedules, but it also provides additional search terms through synonyms. In an online environment a subject-alphabetical index becomes a powerful and dynamic information retrieval tool with no limitation on the number of terms or their hyperlinking. The key feature of any subject-alphabetical index to online classification is its capacity to accommodate new terms that appear either as a result of a new classmark or as the result of feedback based on users' search behaviour (transaction logs). For instance, the search term 'Java' launched against the UDC MRF will only lead to the Java Island and Java Sea and will not retrieve programming languages at *004.434 Individual programming* languages, let alone discriminate between the *Java programming language* and *Java applets*.

This particular example illustrates a typical situation for any general classification: it shows that terms, likely to be sought by users do not exist in the schedules since they are implied under a more general class of concepts or are represented by synonyms. The example also points to a lack of concepts/classes (e.g. *Java applets*) to accommodate new developments in the field - for which there is no specific UDC number, or that the classes that can be used are too general. It is obvious that although the disambiguation of homonyms (*Java place/Java* *programming language*) is inherent to the classification structure, in an online environment, it needs to be 'picked up' by the system through a subject-alphabetical index. To overcome these difficulties a subject-alphabetical index to a classification usually consists of the following operations:

- expansion of terminology by: adding synonyms and quasi synonyms, adding missing terms corresponding to concepts implied in the broader classes, compound and complex concepts corresponding to the compound and complex UDC numbers in the database
- revision of hierarchical relationships: adding terms for missing hierarchical levels and linking terms to the proper broader term in case of invalid hierarchies
- augmentation of semantic linking by: exploiting existing cross references and adding more cross references; gathering distributed relatives; qualifying homonyms

A printed index usually takes one specific form, but when managed through a database, an index can merge several structures. Nevertheless, in relation to its complexity and the support it provides for term finding, we can make a distinction between the following classification index structures suitable for use online:

i) a simple index which links term(s) in natural language to a class that contains concept(s) related to the term(s) in question

ii) a relative index which aggregates all related concepts that are usually scattered throughout the schedules (distributed relatives)

iii) a chain-index, which expresses hierarchical relationships i.e. when it represents concepts in the context of being linked to a broader classes

iv) a thesaurus

(Slavic & Turkulin, 1998)

4.2.2.2.1 Simple, relative and chain subject alphabetical indexes

Simple and relative indexes are very similar but can be distinguished by the fact that a relative index will contain terms that do not exist in the schedule. A relative index will also more consistently collocate distributed relatives, i.e. all the related subjects scattered in different classes, and put them under appropriate headings. A simple index does not make a substantial commitment to identifying semantic relationships and does not have a predominantly syndetic structure.

A word search against the UDC MRF database will often produce the effect of a relative index as it will yield all the occurrences of a given term in the schedules. A true relative index, however, requires careful semantic linking of concepts and takes into consideration the level of specificity to be expressed (i.e. a term may be repeated in a sequence of subordinated classes). For instance, the search term "marriage" occurs 58 times in the UDC MRF database in different contexts but a relative subject alphabetical index is likely to be presented in the following ways (showing only the broad class in a choice of main subject areas):

[Ethics]	173.1	Marriage. Matrimony
[Religion]	2-45	Marriage and the family
[Sociology]	316.811	Marriage and the family
[Law]	347.2	Marriage
[Public administration]	35.083.93	Marriage. Permission or prohibition
[Ethnography]	392.5	Marriage. Wedding. Nuptials

It can be concluded that, in general, a good online subject alphabetical index will strive to exploit semantic relationships in the manner of a relative index.

A chain index is a subject alphabetical index in which each entry is followed by a sequence of superordinated classes and can be offered as an intermediate term dictionary that can help users to decide which term to follow. This approach, advocated and widely popularised by Ranganathan (1962, 1967), helps term disambiguation by bringing the sequence of broader classes under which a term belongs into the subject-alphabetical index. The string of terms represents an edited reverse order of successive class subdivisions. Svenonius, Liu & Subrahmanyam (1991) claim that the importance of a chain index in an online environment is even greater than is the case with manual systems. With the automation of classification systems attempts were made to use computers in assisting the creation of chain indexes.

Figure 4.5 illustrates the first phase in the automatic extraction of a chain index (FATHUM classification), which requires further human control in the selection of entries, the selection of caption hierarchies and the addition of new entry terms. In any automatic extraction of terms from the classification captions it is also necessary to choose the number of hierarchical levels to be displayed next to each term.

Results		
stone	Originals for printing (metals, stone etc - Cartographic images - COMMON AUXILIARIES OF FORM - COMMON AUXILIARIES	(B843-44)
stone	Stone Age - Geological, archaeological and cultural time divisions - COMMON AUXILIARIES OF TIME - COMMON AUXILIARIES	(F631/634)
stone	Rock - Naturally occurring mineral materials - COMMON AUXILIARIES OF MATERIALS - COMMON AUXILIARIES	(L25)
stone	Stone - Naturally occurring mineral materials - COMMON AUXILIARIES OF MATERIALS - COMMON AUXILIARIES	(L25)
stone	Hard-setting compositions - Manufactured mineral-based materials - COMMON AUXILIARIES OF MATERIALS - COMMON AUXILIARIES	(L326)
stone	Imitation stone - Manufactured mineral-based materials - COMMON AUXILIARIES OF MATERIALS - COMMON AUXILIARIES	(L326)
stone	Stone - Gross materials - Materials in art - The arts	940L68
stone	Styles by period - Movements and styles in art - The arts	940P1
stones	Precious stones - Naturally occurring mineral materials - COMMON AUXILIARIES OF MATERIALS - COMMON AUXILIARIES	(L28)
stones	Semiprecious stones - Naturally occurring mineral materials - COMMON AUXILIARIES OF MATERIALS - COMMON AUXILIARIES	(L28)
stones	Other precious and semiprecious stones - Naturally occurring mineral materials - COMMON AUXILIARIES OF MATERIALS - COMMON AUXILIARIES	(L289)
stoneware	Stoneware - Manufactured mineral-based materials - COMMON AUXILIARIES OF MATERIALS - COMMON AUXILIARIES	(L363)
stoneware	Stoneware - Arts and crafts by material - Art forms. Artistic techniques. Artistic media	94011121132

Figure 4.5: Screenshot of automatically extracted chain index in FATHUM classification

A large proportion of terms in UDC, for instance, are likely to have sequence of more than five broader hierarchical levels. Figure 4.6 illustrates how the first step in automatic extraction would look like in UDC.

SUBJECT ALPHABETICAL I.E. CHAIN INDEX (FULL HIERARCHY SPELLED OUT)	
Java	(594.5) Provinces of Java (Java) - Indonesia - Asia - Countries and places of the moden world
Java Sea	(265.742) Java Sea - Seas along the Sunda Islands - South and South West Pacific - Pacific Ocean - Oceans, seas, interconnections, Physiographic designation

UDC SCHEDULES Common Auxiliaries of Place	
(2)	Physiographic designation
(26)	Oceans, seas and interconnections
(265)	Pacific Ocean
(265.7)	South and South-West Pacific
(265.74)	Seas along the Sunda Islands
(265.742)	Java Sea
(4/5)	Countries and places of the modern world
(5)	Asia
(594)	Indonesia. Republic of Indonesia
(594.5)	Provinces of Jawa (Java)
(594.51)	Jakarta Raya
(594.53)	Jawa Barat (West Jawa)
(594.55)	Jawa Tengah (Central Jawa)
(594.57)	Yogyakarta
(594.59)	Jawa Timur (East Jawa)

Figure 4.6: Excerpt of a chain index if automatically derived from UDC schedules

One of the problems in the automation of chain index creation is the addition of terms that are missing in the schedules. In addition, automatic extraction of a chain index requires the 'corrected'[59] and machine-processable data on the class hierarchy. Parameswaran (1990), for instance, reported on the chaining procedure in DDC, highlighted the defects in the DDC hierarchy[60], hidden hierarchical levels and a lack of precision in vocabulary that affected the automation of the process.

Svenonius, Liu & Subrahmanyam (1991) also found that for automation it is important to have the class description normalized in order to be able to extract terms but also to have fully modulated hierarchies, i.e. all the levels of a class hierarchy present in the schedules. Their research was based on a system called DORS (Dewey Online Retrieval System) that included class 700 of the DDC. At the project conclusion, based on automation complications caused by the classification itself, the authors had the following recommendations for the editors of DDC: a) develop better vocabulary control; b) normalize the expression of concepts (i.e. when re-used in combinations, the wording should remain the same); c) attempt to reduce the length of terms where possible; d) redress syntactic anomalies (Svenonius, Liu & Subrahmanyam, 1991: 360).

Research into the automatic extraction of chain indexing from the UDC shows that the difficulties that stem from its inherited similarity to the DDC are still present in the UDC: telescoping of hierarchies and invalid hierarchies (Kumar & Parameswaran, 1999). These are known problems that can be solved by the UDC revision process through carefully drafted normalization of class captions and the correction of invalid hierarchies. As for the lack of hierarchical levels (telescoping), this problem can be alleviated by recording correct hierarchical links in the database. UDC, being an analytico-synthetic classification, may still need more intellectual input in automating the chain index in the actual IR applications.

The issue of pre-composed classmarks is essentially a question of access through terms attached to each element of a single number. The automation of indexes will depend on either more effective database management of pre-composed

[59] As mentioned earlier in this chapter UDC notation does not always express hierarchy, Loth (1996) call the creation of proper hierarchy data 'correction of incorrect notations'.

[60] Among these defects Parameswaran thinks of class numbers representing several concepts (notational homonyms), mixed arrays and telescoping. Telescoping is a term used by Ranganathan and it refers to the situation when a class appears to be coordinated in the notation while the concepts are actually subordinated. This is often applied in classification to shorten the notation (Bhattacharyya & Ranganathan, 1978).

classmarks, or on the decomposition of numbers. But at the same time, even the crudest chain index produced by the automatic extrapolation of a hierarchy based on a permutation of caption terms may prove useful.[61]

4.2.2.2.2 Thesaurus as subject-alphabetical index

Thesaurus extraction or *thesaurification*[62] is the most complex and most ambitious way of creating a subject-alphabetical index for the UDC. As a means of full vocabulary control it has many advantages. Having a subject-alphabetical index in the form of a thesaurus imposes better control over classification maintenance and revision but also allows better use of the classification in information retrieval. The possibility of automating the building of a thesaurus from the UDC by capitalising on the reliability of the classification structure is extremely attractive for users.

Riesthuis & Bliedung (1990, 1991) tested the creation a subject-alphabetical index to the UDC in the form of a thesaurus on class *314 Demography* which was in the process of revision in 1988. The objective was to have a thesaurus as a source of search terms while the system would still use classification for actual retrieval.

In building a thesaurus, the rôle of classification is to provide management of polyhierarchies and to prevent conceptual gaps in vocabulary. The rationale for merging two systems is the complementary manner in which vocabulary control works in thesauri and classification. As Riesthuis & Bliedung explained, while a thesaurus has rigorous control over the linguistic forms of words, it may allow the existence of descriptors that overlap in their meaning. Although terms may be automatically extracted from the caption, they do not always convey the full meaning, and this is apparent only from a hierarchical context; the main intellectual effort is in the actual assignment of a descriptor corresponding to the classmark.

Riesthuis & Bliedung demonstrated that even the part of the schedules that is

---

[61] For instance, a chain index based on a faceted structure has been automatically produced in the FATKS project available at (http://www.ucl.ac.uk/fatks/php/search.php). This can be considered as an initial stage in creation of a fully functional subject-alphabetical index.

[62] The idea of having an indexing retrieval tool in which the classification would be fully integrated with a thesaurus was suggested in the past, such as 'thesaurofacet' compiled by Aitchison (1970), or 'classaurus' as introduced by Bhattacharyya (1982) and Devadason & Kothanda Ramanujam (1982).

still enumerative[63] can easily be translated into descriptors which may be an immediate remedy in IR for the obsolete classes awaiting revision. In this instance, classes which contain enumerated compounds would be described with more than one descriptor and/or would be described by a complex non-descriptor. Then, classes which contain a caption with paraphrases would be represented with a new term, e.g. '*families without children*' into '*childless families*'. There was an obvious need for great deal of semantic factoring in the case of enumeration e.g. when a class description contains meaning inherited from the class above or from several classes. The conclusion of the project was that creating a thesaurus may provide great assistance to the revision process as it gives better insight into the standardization of structure and terminology.

Thesaurus extraction from the UDC could be the final stage of one of the previously described procedures of building an index. Zürich's ETH library system has been noted as an example of good practice in searching classification (Buxton, 1990), and has started a 'subject register' that contains UDC numbers with terms in English, French and German. Based on the UDC schedules they initially produced: an *alphabetical subject register* (as a trilingual unified register); a *systematic register* (three language parts); and a *descriptor environment* for a chosen term in three languages (Hug & Nöthiger, 1988; Hug & Walsner, 1990; Loth & Funk, 1990; Buxton, 1991; Loth, 1996). The system evolved to become subject authority data with UDC numbers fully mapped to what could be called a multilingual UDC thesaurus.

A similar undertaking at the University Library of Bucharest was described by Francu (1996, 2000, 2003a). She first created a simple dictionary of UDC classmarks, but the expansion of the dictionary created the need for better term control (synonyms, homonyms). While her initial research reported in 1996 was on UDC class *8 Linguistics and Literature*, the methodology of extracting a thesaurus from the UDC was later expanded on a general level, with the focus on creating the thesaurus as a multilingual interface in Romanian, English and French (Frâncu, 2000; 2003). She was especially interested in thesaurus extraction from pre-combined

[63] I.e. compound concepts are denoted with a simple notational symbol and are listed in the same hierarchical sequence as simple concepts.

numbers rather than UDC schedules, which required a significant amount of semantic factoring and structural mapping. She also created a multilingual thesaurus to the English UDC Pocket Edition (BSI, 1999) which contained a significant number of pre-combined classes (Frâncu, 2000).

In many instances, descriptors corresponding to simple UDC numbers ought to be complex in nature i.e. some simple UDC class may be represented with more than one descriptor, while some UDC compound and complex numbers can be represented by a single descriptor. She also emphasises the importance of the difference between semantic and structural mapping between an assigned thesaurus and UDC classmarks, and the need for semantic factoring (Frâncu, 2001).

The starting point in connecting numbers and words for the purpose of IR is the differentiation in semantics of simple as opposed to pre-combined UDC numbers, as they appear in both the UDC MRF and bibliographic databases. The problem was addressed earlier by Riesthuis who initially focused on the logical requirements of post-coordinate searching of what is in essence a pre-coordinate index. If verbal expressions corresponding to UDC numbers are to be used in searching, then a link between every single pre-combined element and its description ought to be preserved. In order to search UDC using words, and in the case where the system does not provide access to elements of complex numbers so they cannot be linked to their description, it is necessary to split pre-composed numbers (Riesthuis, 1994, 1999).

4.2.2.2.3 Mapping UDC to subject-alphabetical indexing languages

The motivation for using the classification as a 'pivot' may be the need to expand the scope of some special vocabulary, to improve browsing and navigation functionality of an indexing language already in use, or to provide a common semantic structure to map several indexing systems. Campochiaro & Marcelli (1991), reported on mapping of vocabulary used in a parliamentary library[64] to the UDC structure in order to establish and exploit classificatory semantic relationships in vocabulary control and information retrieval. Himanka & Kauto (1992) reported on 'translating' UDC numbers from the "Finnish abridged edition of the UDC" to the

---

[64] The research was conducted in the Biblioteca del Senato della Repubblica in Rome.

general Finnish subject heading system in an automated fashion. Describing the process, they made a distinction between 'synthesised numbers', 'shallow enumerative' and 'deep enumerative' classes. They experienced no problem in translating synthesised numbers and described the process as smooth.

Similar to Riesthuis, Himanka & Kauto pointed out that there was no problem in mapping enumerative classes to subject heading systems, but they made a distinction between shallow enumerative and deep (specific) enumerative classes. They found classes with captions such as '*other fabric-forming machines*' especially troublesome. These would occur at the end of the hierarchy and they simply did not map them. In the class *677 Textile industry*, when the goal was to translate the complete depth of the hierarchy, they found that 43% of classes could not be translated; in class *539 Physical nature of matter* the result was 45%. Their suggestion was that the bottom of the hierarchy with the most specific classes should be omitted - as a result the success in automatic translation was over 90%. In summary, they found that they could not automatically translate around 16% of the total abridged edition on which they worked. Their objective was to produce machine-readable UDC data that could be straightforwardly implemented into VTLS library systems in Finland.

In present-day library systems, where subject data are usually maintained and managed through authority records, it is possible to hold and establish links gradually between the UDC and any other system in use. The Library of the Croatian Academy of Sciences and Arts in Zagreb, for instance, built their subject authority file so as to integrate UDC with the subject heading systems in English and Croatian (Leščić & Cvitaš, 2003). In 2002-2003 the Czech National Library (CNL) started three projects to improve subject access to the cooperative library catalogue. These included building complex subject authority data that will contain a mapping of the UDC to the DDC, and a mapping of pre-combined UDC numbers to LCSH terms. In addition to this, CNL will connect the translations of the UDC in English to the national subject heading systems of several East European countries (Croatian, Latvian, Lithuanian and Macedonian) in their authority file. In this project UDC is treated as an enumerative classification dealing with pre-combined numbers only (Baliková, 2003).

4.3 Summary of UDC use in IR

The UDC has an undisputable value in indexing and retrieval. The UDC hierarchical structure enables subject browsing from general to specific and its syndetic structure i.e. 'see also' or associative semantic linking between remote subjects allows "lateral" browsing. Also, the UDC's synthetic structure helps to 'fill in' subject gaps and support the indexing of interdisciplinary subjects and allows for their browsing. Findings referred to throughout this chapter show that the hierarchical and synthetic, semi-faceted structure of the UDC can be used for both pre- and post-coordinate searching and browsing. Post-coordinate searching based on the UDC helps achieve subject matches without the user's knowledge of the system while its pre-coordinate indexes preserve subtle differences in meaning and allow logical subject aggregation. The UDC semantic structure allows for multidirectional search expansion, hence, UDC can be used to improving both recall and precision.

The application of the UDC in indexing is scalable. It can be used as a simple taxonomy or for complex number building that allows for great precision in indexing and retrieval. In order to be used to its greatest potential, whatever the complexity of its implementation, the classification has to be interfaced with natural language. There are different options in creating a subject-alphabetical index that have been explored already and common to all of them is that they are based on a classification authority tool which enables term control and linking to UDC numbers.

Some of the requirements for UDC retrieval are general in nature and are the same as in any other classification system. For instance, the support for hierarchy browsing independent of the notation, lateral browsing, searching using words and facet browsing. Other requirements are more UDC specific, these are: control of relationships in complex numbers, full facet coding and management that is then related to the sorting of UDC numbers. Although decomposition of pre-combined numbers, in order to search for individual elements or to link the numbers to verbal expressions, is possible, it appears that authority control of classification data independent from bibliographic surrogates is the only solution that would meet all the needs of IR. Only when this is in place is it possible to create an intelligent interface that enables users to use the classification with ease, search using words or navigate the classification structure freely.

CHAPTER FIVE: INTERFACE TO CLASSIFICATION

5.1 Introduction

Research on subject searching, referred to earlier in Chapters One and Four, shows that there is a variety of information seeking situations which cannot rely on a known, well-defined, specifiable subject topic. In effect, subject searching is often exploratory and bears similarities to the learning process (Bates, 1989; Belkin, 1998; Marchionini, 1995; Hildreth, 1991). Modern IR system design has moved from being output-orientated, to being process-orientated with an emphasis on subject browsing and navigation, which made the rôle of classification more prominent. Visualizing subject access based on classification, has become even more popular with resource discovery on the Internet, as explained in greater depth in Chapter Three. The advantage of using classification in IR, however, depends on the power of the GUI, which enables user-friendly classification browsing and searching. The power of this interface is in supporting visualisation that will 'convert' what is potentially a user-unfriendly indexing language based on symbols, to subject presentation that is easy to understand, search and navigate.

When classification is made available for use in an online environment the same data are used for both creating classification tools to be used by indexers, and for information retrieval to be used by end-users. Equally, an appropriate interface through which the classification is going to be accessed can be suitable for classification tools and for information systems based on classification. This chapter explains some basic functions in browsing and searching that need to be supported in relation to the UDC, irrespective of any specific, real-life implementation.

UDC, being a semi-faceted system can be used for both post-coordinate searching, and hierarchical/facet browsing, as explained earlier in Chapter Four. The versatility of its application is, however, proportional with the sophistication of the IR system and its interface. A summary of basic functions that are expected to be supported on the interface is given in this chapter and the relationship between the flexibility of the interface and the management of classification data are outlined.

The interface of OPACs, as the most widely used IR systems that use classification, is observed in order to find out what is the most likely use of UDC in practice. In this chapter, the number of IR functions in relation to the UDC is recorded and analysed for each OPAC, in order to illustrate the level to which contemporary library systems are able to support and exploit UDC.

5.2 Visualization of subject access

Hancock's study (1987) on subject searching behaviour shows that users tend to browse shelves even if their subject is well understood and defined in advance. According to her research, they tend to combine a catalogue search with subject browsing on the shelf, in spite of the fact that they do not fully comprehend the classification structure. Also, she found that they tend to focus on the immediate subject proximity on the shelf, i.e. a few books before and a few books after. Her conclusion on the importance of collocation and aggregation is probably the most important point in creating an interface to classification as it stresses the importance of dense collocation and very localised but intensive linking.

Linear, pre-sequenced and inflexible browsing of classification schedules that display the subject in an extensive chain of classes and subclasses obviously carries very little importance compared to more localised, interlinked, multidirectional, associative subject displays. These browsing and navigational paths depend on the interlinking of classification data, as well as linking the classification to bibliographic data. The supplied connectivity represents a rudimentary form of hypertext that offers the possibility for users to select and follow a desired link and change direction at any single point of browsing.

While there is a significant correlation between classification browsing and improving recall, research on DDC and LCC shows that navigation through a single hierarchy is not likely to improve precision (Hildreth, 1991; Markey Drabenstott, 1990, 1996). Research on faceted thesauri, however, shows that these are well adapted for search expansion through both hierarchical and associative relationships, provided that semantic linking in the vocabulary is fully exploited (Binding & Tudhope, 2004). Similarly, a semi-faceted and fully synthetic classification, such as UDC, offers parallel facet hierarchies and lateral linking and allows for search results to be further specified or broadened.

An analytico-synthetic classification structure enables precise indexing as well as the visualisation of hierarchical, coordinated and collateral relationships among concepts and, if an interface reflects the structure of the classification, this can help users expand a search, combine terms and specify complex subject components. The need for visualisation of subject access was especially emphasised by the expansion of Web information services and this initiated more intensive research on the topic (cf. Visualising subject access for 21st century information resources, 1998).

5.2.1 Hypertext and non-linear browsing

Hypertext technology allows data structures to be dynamically re-arranged and represented in a non-linear manner. Hypertext consists of nodes, each node being a basic unit of information, e.g. a standalone chunk of text, page or frame, linked to a network of nodes. This network can be represented using *hypergraphs* that allow for the unlimited inclusion of a pictorial and graphic display, termed *hypermedia*. Most importantly, interactive (live) hypertext is capable of storing (authoring) new relationships. This technology allows information presentation that coordinates, within a single view, navigation between a full document, an abstract, thesaurus or classification, which makes it an ideal tool for mounting and displaying bibliographic tools (Garcia Marco & Estaban Navarro, 1996).

Hypermedia represents a platform for the coordination of natural-language retrieval and multi-criteria retrieval using the concrete form of nodes. This is why hypertext is considered to be not only a user-friendly interface but also an alternative and complementary access to information. In such a context, a classification node collocates documents according to their content similarity, and the criterion for this collocation should be obvious from the name of the common concept or from the non-formal textual description explaining the meaning of the class and a list of terms associated with the reference terms.

Aboud et al. (1993) pointed out that the influence between classification and hypertext goes both ways: classification can help hypermedia to be more organized while hypermedia will help classification to become more usable. For instance, classification can be used to restrict the investigation domain. Agosti (1991) emphasised the importance of hypertext in administering the semantic structure of a classification scheme or thesaurus (hyper-concept database), and in storing and handling term structure. It is clear that a hypertext system ought to illustrate and give the transparent appearance of the semantic content of various different documents in a collection, while also establishing necessary links between the collection of documents and the term structure, representing the knowledge structure consisting of some kind of classification.

The most obvious weakness of analytico-synthetic classification is that it may be complex to use. Garcia Marco & Estaban Navarro (1996:89:93) suggested that hypertext can constitute a perfect platform for both the automatic management of classification and a user-friendly retrieval interface and can help: *(a)* to make it easier to access and use pre-coordinate languages through combination of an alphabetical search and relational navigation, *(b)* to connect indexing languages with cognitive maps of users (it enables storing of knowledge and learning paths from different points of view), *(c)* to interconnect dynamically documentary languages, and languages with documents, and *(d)* to design, build and manage links between graphical and alphabetical presentation of a thesaurus and classifications.

The most obvious weakness of hypertext, on the other hand, is that it is not instructive but too multidirectional, which leads to confusion. With respect to this, Garcia Marco & Estaban Navarro (1996: 89-93) suggest that hypertext may profit from an intermediary system based on the classification type of knowledge which will: *(a)* organize perspectives and access hierarchical nodes (top-down versus direct activation), *(b)* focus attention (pruning of hierarchies), and *(c)* help in learning paths.

Hypertext is now ubiquitous technology and is considered to be the norm for all IR systems, whether they are isolated or networked. Its full exploitation in library systems led to a great improvement of the OPAC GUI known as *hypercatalogue*. This also opened up new ways of exploiting classification and classificatory indexing languages for subject browsing in OPACs (Hjerppe, 1990; Gödert, 1995).

5.2.2 Classification-based IR features

Use of classification in an IR system may help to meet the basic requirements which are, according to Hildreth (1991:10), that the "user’s query should never produce zero results and the display of the result should not be assumed to be the end-point of the search process". He relates this statement to the premise that "humans have a greater facility for recognizing things of interest when they are encountered than describing them in advance”.

Classification is known to be useful in improving a user's understanding of a subject and it can be used for the selection of search terms through a subject index, for instructing users on a subject through associative navigation, exploratory browsing, and in displaying results as a launch pad for a further, more contextualised search. For this to function, classification has to be represented in such a way that its visual qualities are maximised in both a linear and non-linear direction (e.g. facet browsing, hierarchy browsing, syndetic structure browsing) while its notational representation and complexity ought to be placed in the background. The system should be represented and searched using words but notation searching should also be provided for professionals and expert users.

A general framework for a search interface, as described by Shneiderman, Byrd & Croft (1997) has to provide four phases of searching: *(a)* formulating the query, *(b)* submitting the search, *(c)* reviewing the results and *d)* a refinement phase formulating the next step. When applied to classification, this may contain the following specific steps:

- positioning at the classification tree to start browsing
  - by launching a word search against the index
  - by top/down hierarchy browsing
- submitting a search against a subject area
- reviewing the classified set of results
- moving to browse a specific classification node in the results hierarchically or laterally, or re-launching a query against a specific class

Browsing strategies involve advanced interface features: dynamic queries, smooth integration between the formulation, action and refinement phases coupled with different displays (over viewing, zooming, filtering and relating retrieved documents). In addition, a clear screen presentation of the concepts and their position in the semantic network are necessary (Marchionini, 1995: 101).

There are three types of user interface: command-driven, menu-driven and GUI. The GUI uses graphics (images, icons) to represent options, functions and tasks, and is the only user interface that provides a satisfactory platform for browsing a classification. The characteristics of a GUI, as summarised by Hildreth (1995) are the following: *(a)* sizeable, moveable windows, *(b)* scroll bars to scan through data and lists, *(c)* pull-down menus and pop-up dialogue boxes with preformatted data entry spaces, *(d)* hot buttons for activating functions and *(e)* point-and-click device-based interaction. Combined with hypertext/hypermedia this creates a highly dynamic and interactive environment.

If we put aside subtle differences in spatial abilities between individual users (cf. Allen B. L., 1998: 69), it could be asserted that classification can represent so-called 'spatial metaphor' which makes it possible to visualise a knowledge space and help orientation or help the user find a familiar reference point from which to move forward. Bates (1989) pointed out that, similar to browsing library shelves, which involves random physical or eye movement, browsing in an online environment should also allow random eye movement which can be reproduced by rapid movement across a large amount of text.

Classification, however, contains linear (hierarchical) and non-linear (syndetic) structures that may need many anchors and pointers for the purpose of navigation, as 'flipping back and forth' may lead to a loss of orientation if the computer screen continues to change rapidly. This is the reason why it was so hard to prove the advantages of classification subject browsing in early command and menu driven interfaces. The solution for browsing a hierarchy was in keeping track of broader classes throughout the screens (Figure 5.1).

SCREEN 1

-03	**Common auxiliaries of materials**
-032	Naturally occurring mineral materials
-033	Manufactured mineral-based materials
-034	Metals
-035	Materials of mainly organic origin
-036	Macromolecular materials
-037	Textiles. Fibres. Yarns. Fabrics. Cloth
-039	Other materials

SCREEN 2

-03	**Common auxiliaries of materials**
-032	**Naturally occurring mineral materials**
-032.1	Air
-032.2	Water (in any of its forms)
-032.3	Carbonaceous and hydrocarbon minerals
-032.4	Ores. Metalliferous minerals
-032.5	Rock. Stone
-032.6	Earths. Nonmetalliferous

SCREEN 3

-03	**Common auxiliaries of materials**
-032	**Naturally occurring mineral materials**
-032.3	**Carbonaceous and hydrocarbon minerals**
-032.31	Natural gas
-032.32	Mineral oil. Petroleum
-032.34	Peat
-032.35	Coal, including lignite and anthracite

SCREEN 4

-03	**Common auxiliaries of materials**
-032	**Naturally occurring mineral materials**
-032.3	**Carbonaceous and hydrocarbon minerals**
-032.38	**Fossil resins and copals**
-032.386	Ozokerite
-032.387	Amber

Figure 5.1: Model of hierarchy browsing using of sequence of screens

Furthermore, analytico-synthetic classifications organize hierarchies into mutually exclusive vocabulary facets and sub-facets. Switching from one facet hierarchy to another in this environment involves several command steps and a knowledge of the classification system.

R. B. Allen (1995, 1996) stressed the importance of interacting and manipulating multiple components of a classification in browsing and suggested that a faceted classification would be best presented with an interface similar to an electronic book or an electronic encyclopaedia. In his opinion, the greatest problem appears to be the complex 'interaction between features of the classification itself and between the classification and a collection'. A simple solution from the GERHARD gateway is shown in Figure 5.2.

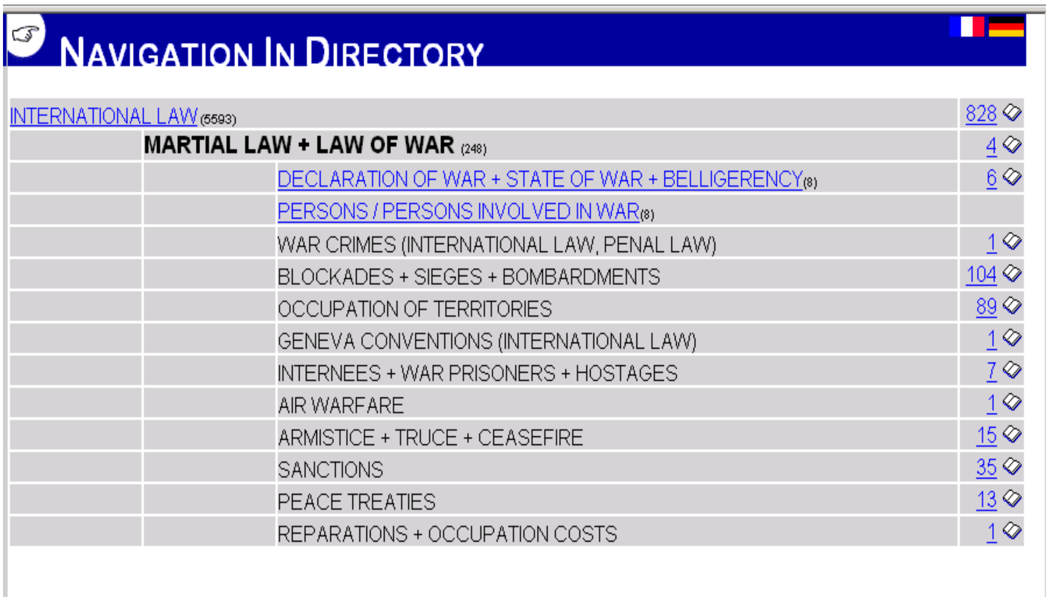


Figure 5.2: Linking classification and collection (GERHARD)

When comparing an enumerative classification interface and a faceted classification structure, R. B. Allen (1996) concludes that the enumerative is easier to present for browsing than a faceted structure. Although not recommended, he points out that the latter may also be reduced to a simple hierarchy. When classification is, however, used for term selection and searching in a GUI environment, in spite of the 'ease of presentation', the hierarchy window may contain an extremely long hierarchy, impossible to view quickly or break down, while in effect the user may want to restrict the search to one single hierarchical node.

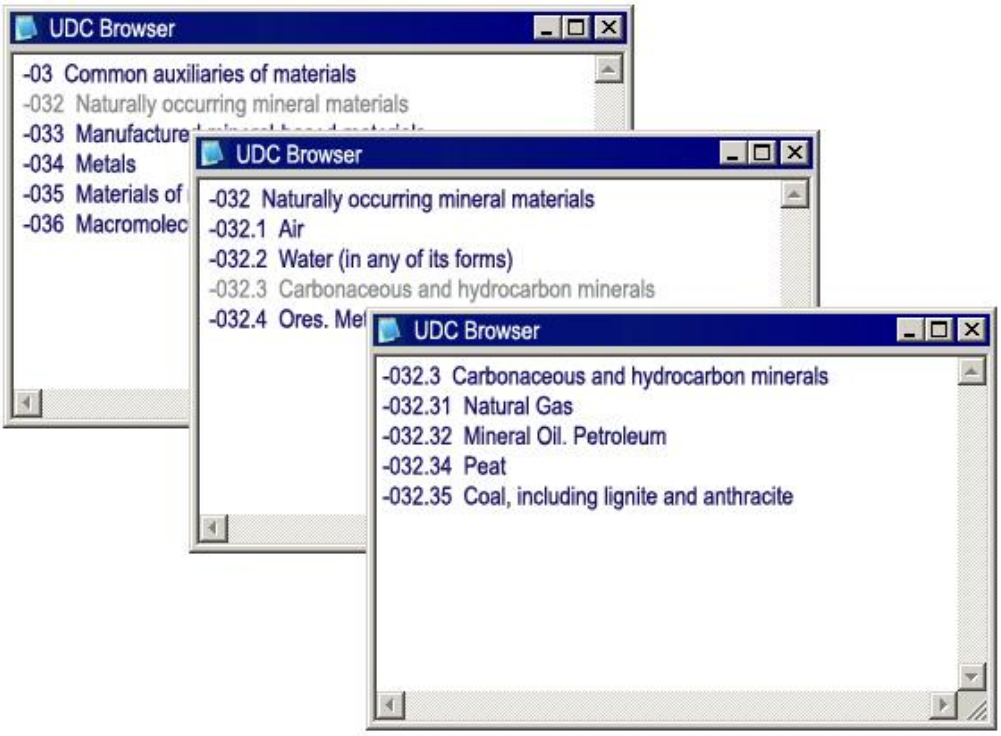


Figure 5.3: Model of hierarchy browsing in cascading windows

Needless to say, GUIs that allow the manipulation of multiple windows have brought a great change to the classification display (Figure 5.3) as they allow an overview of the browsing steps by allowing dynamic and interactive exploration, drilling down the hierarchy and simultaneous opening of several facet hierarchies. The multiple views of hierarchies in separate windows allows the 'starting' hierarchy for browsing to remain present while the link to any subclasses or superclasses is opened as a new view.

The use of interactive windows, each of which can be vertically scrolled, allows viewing of the entire hierarchy while the selection of a hyperlink in one window opens a link in the other. In this case, the main window (frame) can serve as a hierarchical 'menu' while the other frames display a zoomed hierarchical node. The main advantage is that one can have as many windows as necessary, which allows the user visually to separate hierarchical browsing from non-linear (associative).

These interactive window interfaces are obviously more suitable and more exploitable as an interface to analytico-synthetic classifications, which have to be represented as a collection of hierarchical trees rather than a single linear hierarchical sequence.

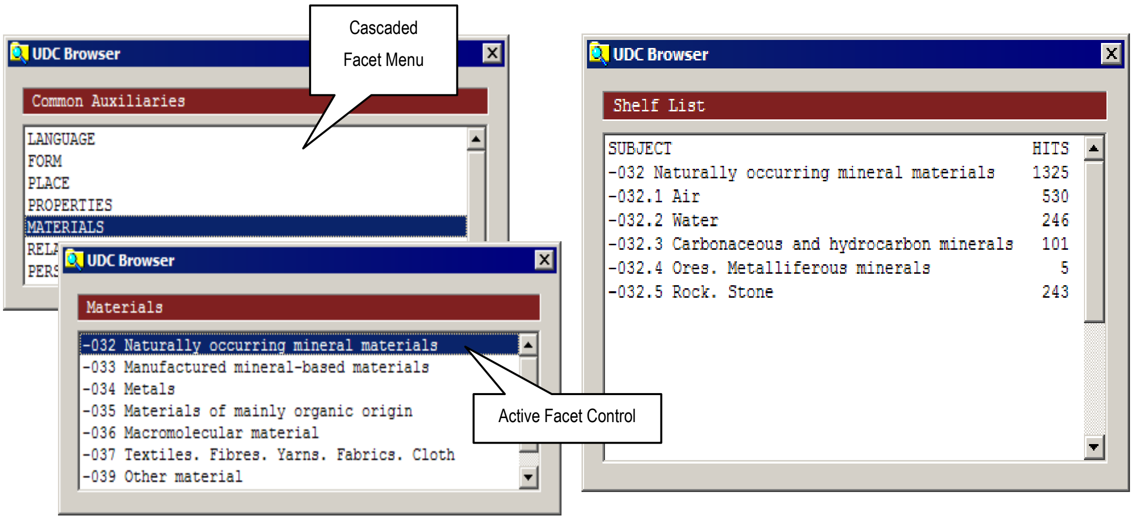


Figure 5.4: Model of cascading interactive windows

The interaction between two or more windows allows better presentation of the complexity of a faceted structure. Typically, as shown in the model (Figure 5.4), interactive window frames would offer the choice of facets ('table of contents') in one window, and a single facet hierarchy in the other, while a third window can be used to zoom into the hierarchical nodes within a single facet. When a collection of documents is classified by a faceted system, an additional layer of complex linking is required, especially since the documents are likely to be linked to several facets simultaneously and may be attached to multiple parents (several facet nodes).

Expanding and collapsing the hierarchy within a single window is an additional function that is important for classifications such as UDC, as it allows for the further integration of a faceted and hierarchical display (Figure 5.5). This is paramount in managing large classification systems that would otherwise require too many windows to be open.

By collapsing all hierarchies, one gets a display of the basic vocabulary facets (top categories), continuing further incremental expansion will gradually reveal the content of facets and the full hierarchy within each subfacet. This feature is probably the most useful in instructing users about the knowledge space and its organization:

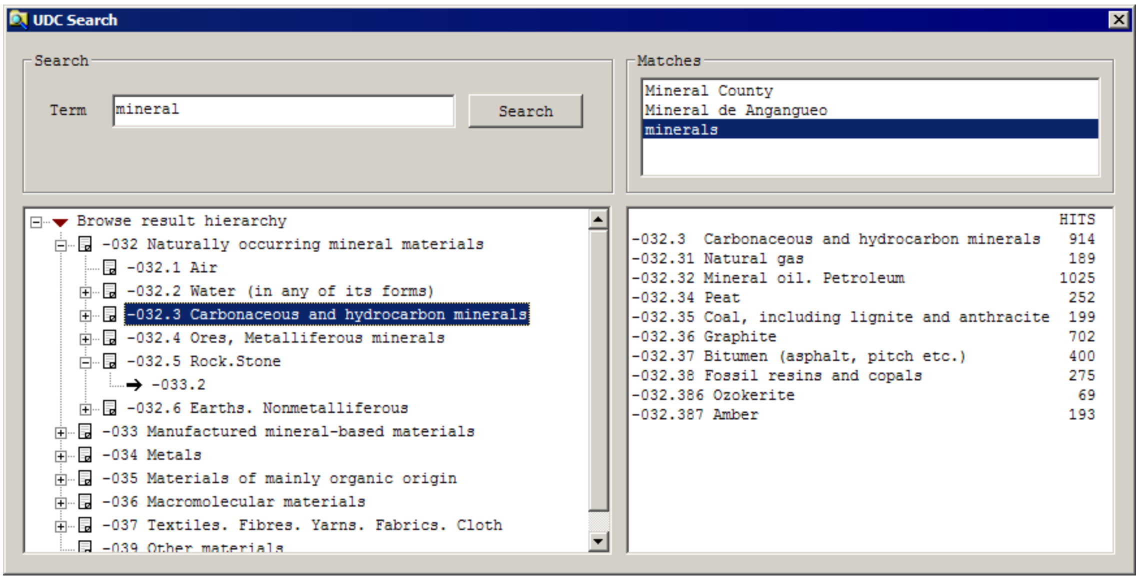


Figure 5.5: Model of browsing interface with a expandable hierarchical list

Independent expansion and collapsing of different facets and hierarchical nodes is very important. In order to prevent a window from overflowing, the opening of a hierarchy node in one place may be set to trigger the closing of all other parent/children displays within the same facet. Icons (buttons) at the hierarchy root are often used to control this function, as illustrated in Figures 5.6 and 5.7 (FATHUM - http://www.ucl.ac.uk/fatks/php/browse.php).

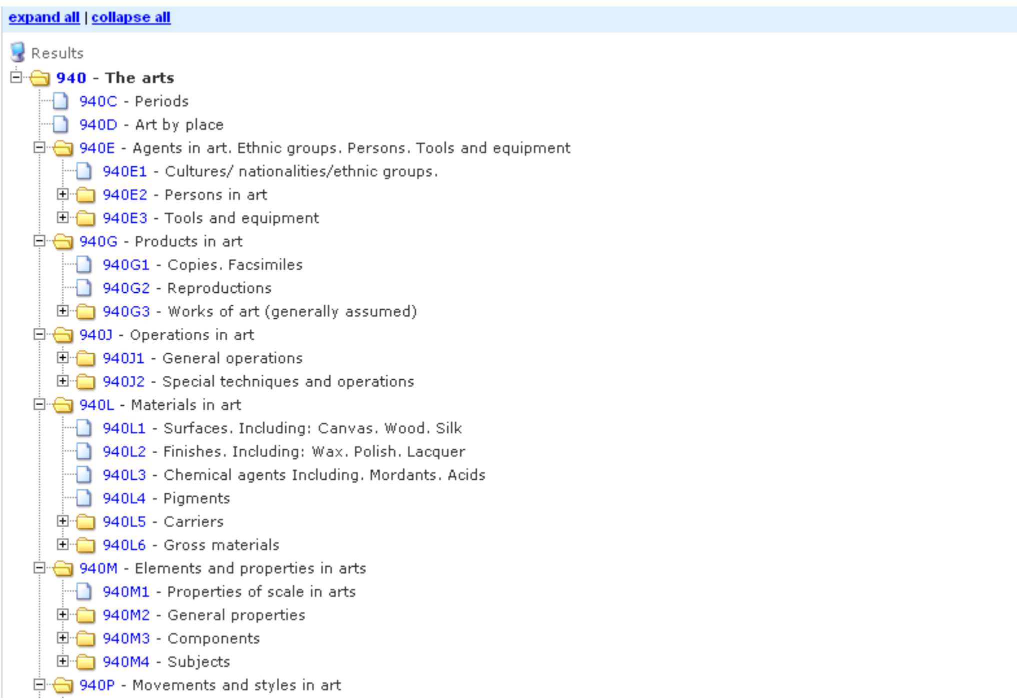


Figure 5.6: Screenshot of tree expansion/collapsing in Web interface

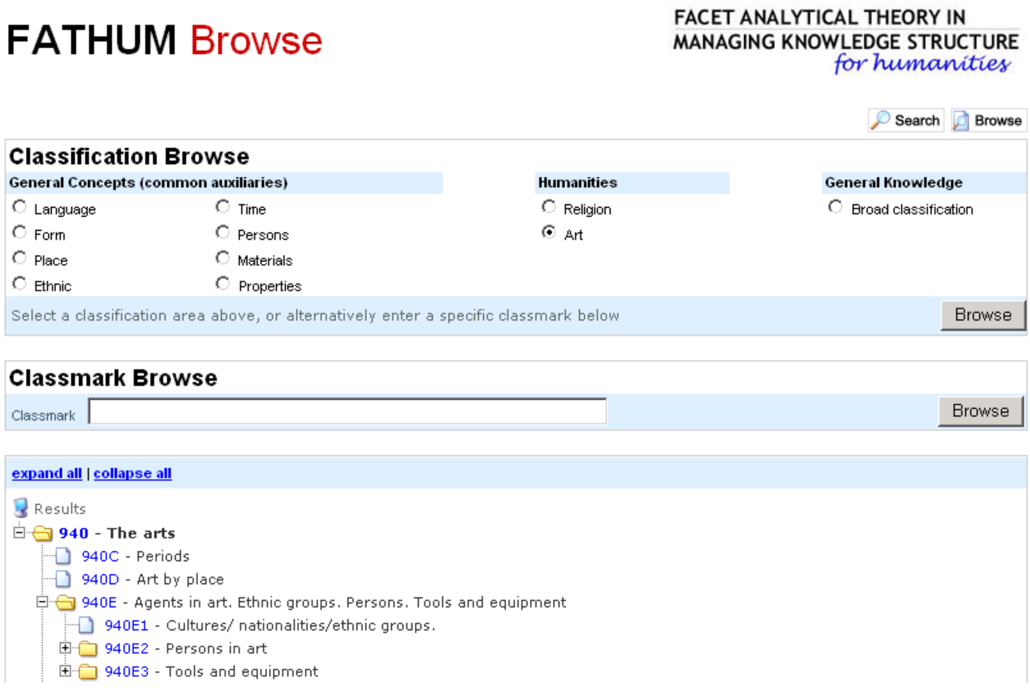


Figure 5.7: Screenshot of facet browsing in a Web interface

The interface for browsing a faceted and analytico-synthetic classification was recognized as being more demanding and more complex to support as it is highly dependent on the possibility of processing classification data automatically (Gödert, 1991a; Priss, 1998; Pollitt, 1997).

Different authors have pointed out that both a hierarchically inexpressive notation and the lack of facet indicators in the notation may influence classification implementation (Pollitt, 1998; Gödert, 1991a; Allen R. B., 1996).[65] Faceted classification may be both broad and deep, in which case the display of facets may overfill the screen. A polyhierarchy on the level of both a classification and document collection is not easy to represent, so in addition to hypertext and interactive window technology there is often a need to include graphic aids in the form of icons (widget buttons), text size and colours. As pointed out by R. B. Allen

[65] In discussing these specific problems in implementation of classification it is useful to make a distinction between problems caused by the lack of machine processable classification data, the inconsistency in indexing or the inconsistency in classification structure (see also the summary of mixed problems by R. B. Allen, 1996).

(1996), colours give an extra dimension to the display and are useful in characterizing the state of the interface and in providing rapid visual indication of selected facets or terms. The use of different sizes and colours of text is yet another helpful feature in browsing a faceted classification and is exploited in many applications.

Research into the implementation of classification browsing on an IR interface was encouraged by a general trend in improving the GUI for OPACs in the 1990s. (Allen R. B., 1995, 1996; Hildreth, 1995). The application, known as *view-based searching*, was studied and demonstrated on a system called HYBROWSE using DDC by Pollitt. This application made use of windowing technologies to provide intuitive and easy to manage browsing, expanding and collapsing of hierarchies. The feature most emphasized in the application was the view of facets used in DDC, i.e. the possibility to view, browse and search standard subdivisions such as a geographic area table (Pollitt, 1997, 1997a, 1998; Pollitt & Tinker, 2000).

The variants of interactive windows are a 'cascading-menu interface' and a 'facet space interface' as described by R. B. Allen (1996). When a faceted classification is applied to a collection of documents, the result is that, in addition to a series of interactive classification hierarchies (cascading facets), there will be a window representing a virtual document shelf as well as a window for term selection (constraints).

A 'facet space interface' is an improvement to a 'cascading-menu interface', achieved through better linking and better control over constraints that prevent facets overflowing and reduce the possibility of exclusion and intersection of terms. The window of the virtual shelf in this case shows only the documents that match these constraints.

5.2.2.1 A Web interface to classification

The Web represents a natural environment for the full use of classificatory structures, as every function that requires a windows-based GUI is easy to construct in this environment. Interactive interfaces on the Web at the moment, however, require customized programming for each website.

The BSI's *UDC-online*, for instance, is a classification tool that is a good example of a Web interface to a classification with facet control using windows and frames where hyperlinks are maximised at every level of browsing, linking frames as well as providing non-linear browsing. To avoid an overwhelming number of windows, colours and graphic aids are used to the fullest extent (Appendix 7.2.1). A different example using the UDC is the interface to the GERHARD search engine, and the directory shows how the UDC can be maximised, preserving simplicity and straightforwardness in a hierarchical display (Appendix 2.7).

Web searching, which is based on the interrogation and accessing of a remote repository may, however, suffer because of the large amount of data that needs to be streamed, rendered and displayed simultaneously by browsers. One consequence is the reduction of speed in browsing and searching.

Relationally linked classification data are quite demanding to display and in combination with interactive windows, hyperlinks and a large collection database, this increases the requirements for display resources. Binding & Tudhope (2004) have reported on this problem in relation to a project that explored IR based on the AAT thesaurus. Each expansion of facets and hierarchies on the Web interface occurs through a new independent page request and download. This causes delay and also, in the search process, the status of the user's query that is linked to the previous page is lost while the current view is being refreshed by new, incoming data. It is not possible to overcome this problem without additional scripting and browser configuration (Binding & Tudhope, 2004).

Marchionini & Brunk (2003) pointed out that there is an increasing need for more standardized tools and services to manage GUI details. One such development is a general class of user interface called a Relation Browser (RB) that allows for the exploration of relationships between different attribute sets. RB manages and controls hypertext, expandable (cascading) hierarchical lists, tabbed views, a zoomable space, magic lenses, mouse-over pop-ups (tooltips), coordinated lists (supermenus) and animations. The greatest advantage of this tool is the display of information by a simple mouse move without the need to click and download data.

Another development, yet to be fully tested for accessing and representing classification data on the Web, is 'programmatic interfaces' within the development of Web Services. Web Services are m2m software applications best described as "a suite of protocols that define how requests and responses between software applications should be encoded (using XML) and transferred (e.g., over the Web using Hypertext Transfer Protocol - HTTP or e-mail) and how such services should be described and registered for discovery and use" (Tennant, 2002). Based on XML technology this service-orientated architecture leads to a separation between the underlying data resources and interface components. OCLC, for instance, research the possibility of 'terminology services' as a type of a Web service that provides mappings from a term in one vocabulary to one or more terms in another vocabulary. Terminology services can be used in the process of cataloguing or information retrieval (Vizine-Goetz, 2003; Vizine-Goetz et al., 2004, 2004a).

5.2.3 Overview of interface functionalities

When discussing the classification interface, Cochrane & Johnson (1996) pointed out that there is a need for better interaction between vocabulary displays, hypertextual browser and retrieval functions and they particularly emphasised the need to support the transition from search to browsing (Cochrane & Johnson, 1996). A poor data structure, however, may impose fundamental limits on the search and interaction options that may be presented at the user interface (Hildreth, 1995).

If a database does not contain information on relationships (hyperlinks) between, for example, a UDC number and its broader class or a UDC number and its caption, or UDC notation and verbal expressions, no interface technology will overcome these limitations. This is why it is necessary not only to pay attention to functions that need to be supported, but also to the data necessary for underpinning those functions.

The requirements for the classification interface are directly linked to the functionality an IR system aims to support. Functions on the interface that are mentioned here should be general considerations for using classification to support searching, browsing and result displays. Since the UDC may be used for both searching and browsing and can be searched using notations and words, these elements should be considered along with specific techniques that may be used to make the most of the existing data.

Based on the Gödert general functional requirements framework for a subject search interface (Gödert, 1996) it is possible to define the presence of these requirements when creating a search interface for the UDC.

As shown in the Table 5.1, most of the functions related to searching are expected irrespective of whether searching involves UDC numbers or their word representation. While Boolean operators and truncation are relevant for number and word searching, adjacency and distance-proximity operators are not very relevant for searching notation but may be relevant in searching captions. The same applies for linguistic intermediation.

Table 5.1: Functions in searching UDC

SEARCHING UDC		
FUNCTIONS	**NOTATIONS**	**WORDS**
Boolean logical operators	●	●
Categorization i.e. broadening and narrowing search	●	●
Truncation/exact match/phrase search	●	●
Adjacency, distance- proximity operators		●
Linguistic intermediation -morphological tools -insertion of the dictionary		 ● ●
Free-text search	●	●

Browsing should be performed on numbers with the caption, or on the caption alone. Browsing numbers with no associated verbal explanation is not a desirable option for end-users. Table 5.2 below shows the interface elements that are important for browsing and transition from searching to browsing or browsing to searching functions.

A term dictionary can be open for alphabetical and systematic browsing and is usually displayed alphabetically but within each entry it may be possible to insert a hierarchical tree. If the term dictionary offered is in the form of a thesaurus this may be further exploited for browsing.

Table 5.2: Functions in browsing and classified display

FUNCTION	INTERFACE ELEMENTS
browsing	• knowledge map navigation window • class numbers with captions display • number of hits against class • hyperlinks on class numbers, see also references • display of targeted class with a broader class and possible narrower classes • different colouring of facets • display with hierarchical indentation of class captions with highlighted position of the relevant class/term in the hierarchy: • collapsing/expanding function key
transition from search to browse and browse to search	• search box available for term search to be launched against class or facet • term dictionary window • hyperlinks (as pointers or as 'trigger' for the search syntax) - results displayed with class number hyperlinked to the broader class - hyperlink to 'see also' related class numbers for lateral browsing and/or search expansion

The functions listed may be required and relevant for all IR systems using UDC, irrespective of whether or not they are isolated or distributed over a network.

The difference between the use of classification in a networked environment or in an isolated system is how the classification data are managed, how the UDC files are linked to surrogate database(s), and how they are built into the system architecture.

How many and exactly which elements of the searching and browsing functions are going to be implemented depend on the specific IR system environment, the size of the database(s) being interrogated and whether UDC is the only, or just one of the indexing languages and indexing techniques used.

Data requirements for supporting interface functions are usually specific to the type of application which is, in the case of UDC, either a classification tool or IR system. Data necessary to support these applications overlap to a large extent as is shown in Table 5.3. Data elements listed may be considered relevant for the majority of interfaces using UDC.

Table 5.3: UDC data related to interface functions

	REQUIRED UDC DATA	CLASSIFICATION TOOL	IR SYSTEM
SEARCHING	notation	●	
	caption	●	●
	UDC number table coding (may appear on display as different colouring of classmarks). Allows for searching against certain table or certain vocabulary facet within table	●	●
	additional codes regulating proper classified display/sorting	●	●
	unique identifiers for the class number (for system handling)	●	
	number of hits against each classmark	●	●
BROWSING	**ALPHABETICAL-SUBJECT INDEX**		
	alphabetical list of terms extracted from caption enriched with missing terms	●	●
	synonyms (hyperlinking of synonyms)	●	●
	qualifiers for homonyms (or broader class description)	●	●
	aggregation of related class numbers (distributed relatives)	●	●
	chain index display or thesaurus display	desirable	desirable
	HIERARCHY & FACETS		
	notation	●	optional
	caption	●	●
	expandable/collapsible list of facets on the menu bar (common auxiliaries and top classes) and option to select one or several. Highlighting of selected classes.	●	●
	opening hierarchical tree in a separate window/frame	●	●
	link to broader class and narrower classes (coded within a system and displayed as hyperlink)	●	●
	link to related classes (see also) (coded within a system and displayed as hyperlink)	●	●

5.3 UDC in Web OPACs

Classification has always been and still is the most common indexing language in libraries, and these are the natural environment to look for a classification performance in IR. Paradoxically, OPACs happen to be an area of application where classification is not used to its full potential. In the 1990s, with the development of library systems using the GUI and later Web interfaces, the conditions for exploiting classification improved.

In the beginning, Web OPACs offered only a part of the search and browsing functionality that was available in the library system interface on-site, but this has gradually changed and the Web has provided the main search interface in many library systems. New generations of hypertext-enabled OPACs have much more potential in visualising relationships, linking and generally in supporting user-system interaction. According to Markey Drabenstott & Weller (1996, 1996a) new OPACs are supposed to address more demanding patrons and are gradually moving to fulfil more requirements for subject searching that were ignored by library catalogues for decades.

5.3.1 Classification in library OPACs: background

Gödert & Horny (1990: 66) noted that a number of user studies in the USA confirmed that subject search represented, on average, 51% of the total searches. Gödert & Horny, however, found out that this was the area where OPACs scored extremely poorly by failing to provide the necessary recall or precision. The authors summarised the IR issues in OPACs as being related either to problems in finding the correct search terms, or to the problem of increasing recall, or the problem of increasing precision. Research undertaken by Markey Drabenstott (1996a) also confirmed that subject retrieval in OPACs resulted in both too great a recall i.e. high posted searches or too small a recall i.e. low posted searches.

Research provided at a very early stage by Wajenberg (1983), Svenonius (1983), Cochrane & Markey (1985), Markey Drabenstott (1986, 1990) emphasised the advantage of using classification in library OPACs for improving recall and precision. Related research by Hildreth (1991) and Markey Drabenstott (1996) indicated that classification can be a device to find additional retrieval in low posted searches and can be useful for relevance feedback. These findings, however, did not lead immediately to better exploitation of classification, as was made evident from later reports in the 1990s by LeBlanc (1995), Marker (1996), Markey Drabenstott (1996, 1996a), Markey Drabenstott, Burman & Weller (1996) and Markey Drabenstott & Weller (1996, 1996a).

The difficulty users experienced in using OPACs, reported in the early eighties, it seems, made no impact on libraries and library system vendors (cf. Howard, 1990; Kniesner & Willman, 1995). Commercial library systems did not support classification as a tool for subject access and this remained the problem throughout the 1990s (Marker, 1996; Cochrane & Johnson, 1996; Markey Drabenstott, 1996).[66] Librarians, for instance, were sometimes reluctant to discard their card shelflist as library systems did not provide the necessary functionality in sorting and browsing classification or control over call number assignment.

Analysis of eight widely used library systems in 1995 by Kniesner & Willman (1995) showed that only two of them were able to support basic requirements such as: shelflist order, enough space to display the full classmark, display of the main entry, free browsing (forward/backward), and approximate matches. Even shelflist browsing as introduced in the latest generation of OPACs, especially Web OPACs, is quite restricted for no good reason (LeBlanc, 1995; Hildreth, 1991; 1995; 1995a).

When analysing a selection of bibliographic databases and in-house library systems using UDC in 1990, Buxton illustrated a sophistication in classification use that was not common for library systems in general (Buxton, 1990). The selected libraries developed innovative solutions in order to search and browse UDC.

Libraries using vendor systems, however, such as the Cranfield Institute of Technology Library (Libertas), Aberdeen University (Dynix), Edinburgh University (Geac), and King's College London (Libertas) are mentioned as examples of catalogues that had no special features for handling UDC numbers (Buxton, 1990:

---

[66] "To spend enormous amounts of money assembling book collections and then to make access to them only superficial, partial, incomplete, and haphazard is to throw money away rather than to use it prudently" (from Thomas Mann's "Cataloging and Classification Quality at the Library of Congress" cited by LeBlanc, 1995; 301).

200-201).[67]

Other published reports on use of UDC in library OPACs at the time, such as those by Porter & Galpin (1988) on the Muscat system at the Scott Polar Research Institute Library, or those on the ETH Library system ETHICS by Hug & Nöthiger (1988), Loth & Funk (1990) and Loth (1996) were success stories. Throughout the 1990s there were very few published reports on how UDC is searched or browsed in library OPACs.

5.3.2 Study on UDC in Web OPACS in 2004

In order to establish to what extent OPACs using UDC support classification, a small study on Web catalogue interface functionalities was conducted in 2004 as part of this research. Thirty Web OPACs from different countries were observed with the objective of establishing the number of subject searching/browsing functionalities, based on UDC, that were supported. The purpose of this study was to get a general picture on how different libraries exploit the existing potential of the UDC system, library systems and the Web Interface.

The aim in this research was to describe the general level of libraries' willingness and expertise in making use of UDC in providing subject access, rather than analysing any given library system and indexing practice or real users' needs, From this point of view it was relevant to look into the performance of Web OPACs based on the same library system in different countries and with a different scope of UDC usage.

The international selection of OPACs was based on the type of UDC use in a given country. The majority (23) of OPACs chosen were union catalogues in academic, public or national library networks in 22 countries that could be categorized according to the UDC usage as shown in Table 5.4:

[67] The Situation in 2004 is as follows: King's College Library uses Aleph and UDC can be only seen when browsing call numbers; Cranfield Institute of Technology Library (since 1993 Cranfield University) uses the Unicorn Sirsi system and does not provide any access to classification search or browse and UDC can be seen only in call numbers in record display. Aberdeen University Library uses Aleph and provides classification browse to DDC and no UDC is visible in the catalogue. Edinburgh University uses Voyager and does not provide any access to classification search/browse, only the Main Library seems to be still using UDC for shelf arrangement and this is only visible through record viewing. Imperial College London uses Unicorn and also does provide no access to UDC search or browse.

Table 5.4: Selection of OPACs according to type of use of UDC

USAGE CATEGORY		COUNTRIES	COUNTRIES IN CATEGORY	OPACS IN CATEGORY
A	UDC is used in majority of libraries	Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Portugal, Slovenia, Macedonia, Serbia & Monte Negro[1], Spain	11	13
B	UDC is used in certain type of libraries	Australia, Belgium, Finland, Germany, Ireland, Italy, Norway, Switzerland, U.K.	9	14
C	UDC is used seldom	Austria, Israel	2	3
TOTAL			**22**	**30**

A larger proportion of OPACs was chosen from countries in groups A and B, where UDC was used in the majority of libraries, and from those countries where it is widely used in academic and special libraries, as these countries are more likely to have a history of implementing the UDC in library networks and more incentive to make full use of the classification. Care was taken that catalogues represented both systems developed in-house and well known commercial products.

The assumption was that in-house library systems impose fewer limitations in developing appropriate tools for classification management and searching. Of all these Web OPACs, 5 were in-house and the remaining 25 were one of the following vendor systems: Aleph (8), Athena, EOS Glas (1), Horizon (2), Innopac (2), Bookmarc Sirius (1), Talis (1), Unicorn Sirsi (3), Voyager (1) and VTLS (2).

Web OPACs were selected based on a combination of a random search for library OPACs in countries were UDC is known to be widely used[68] and literature research. The number of OPACs was chosen based on literature research (Buxton, 1990; Baliková, 2001; Gödert, 2003; Gobytė, 2000; Leščić & Cvitaš, 2003; Hajdu-Barat, 2004, Schallier, 2004, 2004a; Gnoli, Ridi & Visintin, 2004).

While some were chosen based on the fact that they were more advanced in their use of the UDC others, however, were selected as they best illustrated the situation in the respective environment.[69] OPACs listed alphabetically by country are

[68] In the period from 2000 to 2004 a number of Web OPACs addresses was gradually collected through library checking and correspondence with UDC users and published on the UDC Consortium Web pages (http://www.udcc.org/opacs.htm).
[69] Such a case is the choice of three catalogues in Croatia that show discrepancy in functionalities although they are in-house systems specially designed to meet the needs in the environment in which UDC is used as the only classification for both shelf arrangement and detailed subject indexing. Another example is the choice of four UK OPACs, none of which offers suitable functionality for searching and browsing UDC.

given in Appendix 4, Table 4.1 with URL address, type of catalogue (single, union, type of library system used (I for in-house; V vendor) and name of the system. Interface language(s) are indicated and each OPAC is assigned a code that is used in data collection.

5.3.2.1 IR functions observed

The main strength of Web OPACs is the graphic user interface and enhanced hypertext navigation capability. But this does not solve the problems of managing data content, labelling and deciding how these will be grouped, linked, used and presented to the user (Gödert, 1991; Hildreth, 1995, 1995a). Because of this, the attention was focused on 23 pieces of IR interface functionality that are known to be used and are relevant for UDC. In data collection these are grouped into four categories:

- **General interface and library system functions**: type of access to UDC number (search index/call number, browse index/call number); type of UDC use (call number or subject index), availability of list of UDC numbers or words for term selection, use of UDC alongside another classification system; use of UDC alongside a subject heading system
- **UDC browsing**: the existence of a subject directory for top/down browsing; possibility to start browsing through searching (i.e. positioning in the UDC hierarchy); availability of free browsing (i.e. forward/backward), hyperlinking of UDC number in the bibliographic record; the possibility of expanding the UDC hierarchy in the search result (i.e. to broaden or narrow the search)
- **UDC searching**: automatic right truncation[70] in a search (to improve recall); approximate matching (i.e. 'no zero result' option); availability of a Boolean search of the UDC; searching parts of a UDC complex number; searching a UDC caption; searching/browsing from an authority record choosing any of related terms within the record

[70] Left truncation was found in one OPAC (HR-KGZ) (See Appendix 4.2, Figure 4.1). Since this was an exception among Web OPACs this function was not included in data collection.

- **UDC in results display**: classified results display (i.e. collocation of records according to UDC classes); shelflist display (simple list of all items available in the class), correct UDC sorting; UDC caption display (in result or in record), hierarchical indentation of classes in display

A number of functions can be implemented in OPACs based on the presence of the UDC number as a text string in a bibliographical record (e.g. field **6xx** in MARC formats). This means that it does not require special effort for a library to enable these functions on the Web catalogue interface providing it uses any vendor integrated library system. The functions shown in Table 5.5 can only be enabled on the search interface with the management of UDC through an authority control. All library vendor systems that appear in this study are known to support authority control management.

Table 5.5: Relations of IR functions and authority control

	FUNCTIONS THAT CAN BE BASED ON UDC DATA BIBLIOGRAPHICAL RECORDS ONLY (TEXT STRINGS)	FUNCTIONS THAT REQUIRE UDC MANAGEMENT AND ACCESS THROUGH AUTHORITY CONTROL
GENERAL	browse/search UDC index or call number use of other classification systems use of subject heading systems	mapping to other classification systems mapping to subject heading system
BROWSING	search to browse free browsing of UDC index (backward/forward) hyperlink in bib. record display	subject directory/hierarchy browse bt/nt expansion of UDC in search results (irrespective notation) launching related search/browse from authority data
SEARCHING	number search approximate matching (non-zero result)	searching part of complex UDC number (Booleans) searching UDC captions or searching/browsing UDC using words searching UDC using subject heading systems
DISPLAY	approximate classified display approximate shelflist display	correct classified display correct shelflist display UDC caption display hierarchical display

Complete data on the presence of UDC IR functions in all 30 OPACs are shown in Appendix 4.1, Table 4.2.

5.3.3 Summary of findings on UDC in Web OPACs

If examined individually catalogues differ greatly according to their choice of functions and access points to the UDC. It was clear from the outset of this observation that no catalogue needs all the searching/browsing functions or access points suggested that were checked. It is worth mentioning, however, that the number of functions varies from 2 (HR-NL) to 16 (CH-ETH) with 87% of OPACs having between 7-and 16 UDC related access points and functions enabled.

As shown in Table 5.7 on the next page, direct access to searching or browsing a UDC index is offered in 90% of the catalogues. It should be noted that the relevant function can be introduced in different ways on the interface: searching/browsing 'UDC' or 'local classification', 'classification' or 'call number'.

Very few (6.7%) Web interfaces offer an additional expert search, probably because the Web OPAC is considered to be primarily an end-user tool while an expert search is likely to be provided on the library intranet. It is significant that 50% of catalogues are using UDC alongside a subject heading system without evidence that these are actually mapped to the UDC through a subject authority file. In 36.7% cases, however, a UDC caption is used for an alphabetical subject search or as a basis for local subject heading systems (for instance: CH-ETH, BE-UL, HR-KGZ, HR-HAZU).

A subject directory to allow browsing from top-level categories is an unusual function in OPACs and is provided in only 13.3% of catalogues. Three basic requirements are present in a significant number of catalogues: *search to browse*, *free forward/backward browsing* and *hyperlinking UDC from records* (70%, 56.7% and 46.7% respectively). The more demanding function of broadening and narrowing search results by expanding the UDC hierarchy, depends upon coding BT/NT links in the classification authority data and is supported in 23.3% catalogues only.

Two search functions *default (right) truncation* and *approximate match* that can be implemented without maintaining UDC in the authority file are present in 66.7% and 46.7%. A Boolean search of the UDC index is given in only 10% of the catalogues, most probably because this option makes sense only if *searching parts of a complex UDC number* is supported, which is true for 23% of OPACs.

Table 5.6: Summary statistics for presence of IR functions in Web OPACs

	UDC IR FUNCTIONS	22 OPACS WITH UDC IN SUBJECT INDEXING	8 OPACS WITH UDC IN CALL NUMBER	30 OPACs (100%)
GENERAL	ACCESS TO BROWSE/SEARCH UDC INDEX	20 (66.7%)	7 (23.3%)	27 (90%)
GENERAL	NO ACCESS TO SEARCHING /BROWSING UDC	2 (6.7%)	1 (3.3%)	3[71] (10%)
GENERAL	EXPERT SEARCH	2 (6.7%)	0 (0%)	2 (6.7%)
GENERAL	LIST TO SELECT SEARCH TERMS (UDC NUMBERS OR CAPTION)	8 (26.7%)	0 (0%)	8 (26.7%)
GENERAL	UDC USED ALONGSIDE OTHER CLASSIFICATION	3 (10%)	0 (0%)	3 (10%)
GENERAL	UDC USED ALONGSIDE SUBJECT HEADING SYSTEMS	13 (43.3%)	2 (6.7%)	15 (50%)
BROWSING	SUBJECT DIRECTORY FOR BROWSING	4 (13.3%)	0 (0%)	4 (13.3%)
BROWSING	SEARCH TO BROWSE	14 (46.7%)	7 (23.3%)	21 (70%)
BROWSING	FREE BROWSE UDC [FIELD INDEX]	12 (40%)	5 (16.7%)	17 (56.7%)
BROWSING	UDC HYPERLINK [IN BIB. RECORD]	10 (33.3%)	4 (13.3%)	14 (46.7%)
BROWSING	BT/NT BROWSE UDC IN HIT LIST	3 (10%)	0 (0%)	7 (23.3%)
SEARCHING	DEFAULT TRUNCATION	13 (43.3%)	7 (23.3%)	20 (66.7%)
SEARCHING	APPROXIMATE RESULTS (NO ZERO RESULTS)	10 (33.3%)	4 (13.3%)	14 (46.7%)
SEARCHING	BOOLEAN SEARCH OF UDC INDEX	3 (10%)	0 (0%)	3 (10%)
SEARCHING	SEARCHING PARTS OF COMPLEX UDC	7 (23.3%)	0 (0%)	7 (23.3%)
SEARCHING	SEARCHING UDC CAPTION	11 (36.7%)	0 (0%)	11 (36.7%)
SEARCHING	LAUNCH SEARCH FROM AUTHORITY FILE BROWSE	4 (13.3%)	1 (3.3%)	5 (16.7%)
HIT DISPLAY	CLASSIFIED DISPLAY	10 (33.3%)	1 (3.3%)	11 (36.7%)
HIT DISPLAY	SHELFLIST DISPLAY (EACH ITEM WITHIN CLASS)	2 (6.7%)	7 (23.3%)	9 (30%)
HIT DISPLAY	CORRECT SORT OF SUBCLASSES	1 (3.3%)	1 (3.3%)	2 (6.7%)
HIT DISPLAY	UDC CAPTION DISPLAY	4 (13.3%)	0 (0%)	4 (13.3%)
HIT DISPLAY	HIERARCHICAL INDENTATION OF HITS	2 (6.7%)	0 (0%)	2 (6.7%)

In order to allow searching parts of the UDC, some catalogues keep the elements of complex numbers (often only the auxiliaries of place) in separate fields but this was not counted as an actual 'complex number search' function. Only two catalogues (CH-ETH and CZ-NL) have their searches based on authority files that seem to provide complete factoring of UDC complex numbers. Five OPACs (16.7%) allow searching of the UDC from an authority file i.e. the transition from browse to search, in which instance the hyperlink launches a hidden search expression. The

---

[71] In these three OPACs UDC can be accessed through record viewing only. Search has to be performed through some other access points (title, author, subject).

functions of *searching UDC caption*, as well as *launch search from authority file* are related to maintaining classification data in an authority file and these two capabilities are not widely seen within the catalogues examined (36.7% and 16.7% respectively).

In the display of results, again, functions that do not directly depend on authority data are more common in the selected catalogues. Hence, *classified display* (showing classes against the number of records) is present in 36% and *shelf-list* display (showing all items that contain that classmark in the call number) in 30% of catalogues. Correct sorting of UDC numbers still remains the problem, as OPACs seems to continue sorting UDC as text (only two i.e. 6.7% show correct sorting was found).

Only one (3.3%) of the OPACs that use complex UDC numbers that include symbols seems to do this correctly, the other OPAC (with UDC notation in call numbers) was found to have the correct sorting but only because the numbers used did not include any symbols that would disturb the order. Similarly, the presence of *UDC caption* (only 13. 3%) and *hierarchical indentations of results* (only 6.7%) requires the management of UDC in an authority file and it is no surprise that this is not widespread among catalogues.

The assumption that UDC may be well exploited for subject searching even if it is used in shelf arrangement does not seem to be confirmed through observation of 8 OPACs using UDC in call numbers, covered by this study. In general, these OPACs did not show much affinity to including UDC data in a subject search.

It was logical to make the assumption that in-house systems, specially designed to meet the needs of a UDC-orientated library environment, would show better functionality in searching and browsing the UDC. This turned out to be unfounded and the poorest OPAC in terms of functionality proved to be that of the Croatian National Library union catalogue (in-house system) in spite of the fact that all libraries in Croatia are using UDC as the main and often the only indexing language and all the bibliographic data in Croatia contains UDC. Almost as poor was the Slovenian library system that was recently bought by Bosnia & Herzegovina, Macedonia and Serbia & Montenegro library networks. All three countries are using UDC across the library sector as the main and often the only indexing system. The best OPAC in this category is from the public library network in Zagreb (HR-KGZ), Croatia.

The overall impression that transpires from this study is that success in implementing the UDC largely depends on library policy and expertise or need, and not on the library system in question. The functions that do not need special management of classification data are more prevalent across catalogues than those dependent on classification authority data. OPACs using the same library system, such as those using Aleph, show various levels of sophistication. For instance, one example of excellence for providing full functionality of searching the UDC through natural language terms in three languages while retaining all the necessary semantic linking through authority data was the ETH library (using Aleph) as was detailed in Buxton's report during the 1990s. The University of Leuven Library that excelled in using UDC with Dobis Libis, is now reported to be migrating to Aleph with most of the functionalities of the existing classification authority file being transferred and further enhanced as shown in the Appendix 4, Figure 4.2 (Schallier, 2004a).

The link between searching using natural language while displaying results in a classified way which would enable users to determine a relevant group of hits easily is not sufficiently exploited in library OPACs. Hierarchical display of subjects and expansion of a search ought to be supported through proper UDC management and additional hierarchy coding (cf. Loth, 1996). Displaying a caption next to hyperlinked UDC numbers in records and in a classified display could help users to decide more quickly on relevance or may trigger further exploration, and this is often lacking. All the above options closely depend on the separation of IR functions from simple UDC number processing and require full classification data management.

5.4 Summary of UDC interface issues

The ease of use of classification in indexing and its efficiency in retrieval are closely related to how the knowledge structure and its semantic linking are accessed and presented on the interface. Not until the advent of the GUI and hypertext/hypermedia was it possible to consider a more sophisticated interface to classification with dynamic cascading windows and hypertext.

A desirable interface to classification would make use of hypertext and would consist of interactive/dynamic windows that would enable independent browsing of vocabulary facets. It would also allow for hierarchies to be collapsed and expanded, and contain hyperlinks to related classes. Ideally the classification tree would be presented with captions (with or without notation) and appropriate indentation. If the classification interface is created for an IR system, the number of hits should be displayed against each class that is hyperlinked either to the list of surrogates or the list of resources. Classification should be searched using words and a term index may be used to help choose appropriate terms. A transition from search to browse/browse to search should be provided as well as the possibility to scroll back and forward through the hierarchy. Broadening and narrowing of search results should be supported independently of the UDC decimal notation.

The interface functionality is closely dependent not only on the appropriate formalisation and coding of classification data, but also on their enrichment. When classification is managed as a database tool the functions necessary for classification management and indexing and those necessary for IR retrieval overlap. In this context the same classification data are used to support an interface for both librarians (for indexing) and patrons for searching the collection.

The OPAC study in 2004 reported in this chapter shows that the subject browsing and searching functionality of a Web OPAC depends on what was previously present in local systems in terms of classification data. In the beginning Web OPACs were demonstrating only a fraction of the available search functionality on a local level. Gradually, with the Web, they made good use of hypertext technology and they have become more sophisticated and more user-friendly. At this stage of development in library OPACs and Web technology, no discrepancy should be assumed between the usual capacity of any one library system in supporting subject search and its realisation on the Web interface. Behind this application there is both infrastructure and expertise to exploit classification to its full potential.

The study of 30 OPACs illustrates that the vendor's Web OPAC offers a similar package of functionality. At present and with respect to the way UDC data are made available, it is up to libraries to make the necessary investment and enrich UDC with additional data that will make it easier to implement and to 'tune' existing vendor systems. When this is the case, the results can be impressive (e.g. NEBIS system). Other OPACs rarely allow the searching of UDC numbers with anything other than simple truncation. Also, top/down subject browsing, starting from a discipline level, is very seldom present in the catalogues. In-house systems, in this research, show no significant improvement compared to off the shelf, vendor systems.

This OPAC observational study is a small preliminary research that can provide a general impression and initial data for further in depth studies. Compared to the observations of Buxton (1990), Howard (1990) and Kniesner & Willman (1995), it transpires that the situation with respect to vendors' systems is much better than it was a decade ago. Further research and, to a certain extent, existing results in Web OPAC interface research, can be open to interpretation in different ways. If examining how much of the existing potential of library systems is actually exploited, the conclusion may be that there is still a long way to go for the majority of OPACs examined in this study.

When evaluating libraries' willingness and openness towards using UDC in subject searching, one has to take into account the work necessary to turn UDC MRF data into a successful IR tool i.e. the work necessary to create a classification authority file with structured and coded UDC data, rich subject-alphabetical index, coded hierarchical relationships etc. Obviously, if the UDC were distributed as a proper classification tool compatible with library systems and not only as raw classification data in a proprietary format created for traditional publishing, this would facilitate its better exploitation in library OPACs.

CHAPTER SIX: UDC IMPLEMENTATION AND AUTHORITY CONTROL

6.1 Introduction

Most research on UDC use online before the 1990s was concerned with the problems of handling, sorting and searching UDC numbers as text. Decades of application-specific research focused on localised solutions and resulted in very few common technological solutions that could be re-used or shared easily. This chapter focuses on the issues of UDC automation that are still relevant, in spite of the fact that any UDC implementation can be based on a structured data format, as made available through the UDC MRF.

The chapter notes the requirements for the processing of UDC data and covers issues of appropriate management and exchange of classification data via an authority file. Once created, a UDC authority file may be re-used in the process of document indexing, cataloguing and IR, provided that each UDC number in the file can be automatically linked to one or more document surrogates.

An authority file is an important instrument for controlling indexing policy and for reducing the amount of synthesised classmark variants used to denote the same topic. It is highly desirable to hold the UDC in the form of authorized data, for the classification of documents and information retrieval, in such a way that data may be re-used within the same system, as well as shared and exchanged between different systems. This chapter looks into the suitability of existing MARC classification authority formats that may be used as carriers of UDC data.

Successful exploitation of analytico-synthetic indexing languages such as UDC is even more dependent on centralized (authority) control as this is bound to simplify the use of classification, helping validation and re-use of often very complex numbers as well as facilitating their searching and browsing. A UDC authority file in a standardized MARC format can be used in facilitating subject access to distributed library collections.

6.2 Requirements for UDC data processing

When discussing the requirements for the use of UDC in IR, Buxton (1990), found that the most basic are those related to the handling of pre-combined UDC notation syntax. He pointed out that handling of UDC notation as a simple string is a waste of the intelligence that is built into the classification structure and a waste of time and effort for the classifier. This chapter builds further on Buxton's research and, based on what has already been established in Chapters Four and Five, explains issues and requirements in processing three parts of UDC data:

- UDC notation
- hierarchical and associative relationships
- relationships between UDC numbers and verbal expressions

6.2.1 UDC notation

Full automation of the UDC should allow computer programs to: *(a)* differentiate between single and composed notation; *(b)* discriminate and identify each element of a synthesised notation according; *(c)* control the sequence of elements according to specified rules; *(d)* control the display of UDC notation and *(e)* identify and control relationships between notations. The purpose of this is to ensure that the structure of synthesised notations can be recognized and processed by machines for both notation *management* (tracing, validating, global changing) and *searching*. Composed notation can be split and their parts accessed individually, combined and re-combined according to rules processed by machines, which both eases the use of UDC by humans and promotes better usability in knowledge organization and information retrieval.

Based on Buxton's (1990: 199-200) summary of requirements for the use of UDC in IR, it is possible to formulate functions that will underpin both searching of UDC numbers and the full automation of UDC number management in terms of filing, controlling, changing and displaying:

1) processing and searching complete pre-combined UDC notation in a string that may include numbers, punctuation, and symbols.
2) the ability to process and search each component part separately:
   a) truncation of search
      - truncating the suffix of terms in a search e.g. *005.42$* or *005.42(410)-024.6$*
      - truncating in the middle of an expression e.g. *005.42(4$-024.6*
   b) process and search for:

- each notational component in expressions connected with : (colon), :: (double colon) and + e.g. search for *005.81* and retrieve the combinations in which the number is found such as *005.961:005.81* or *005.961::005.336.1* or *005.42(410)-024.6*
- process and search for main auxiliaries in combinations such as *(410)* in *005.42(410)-024.6*
- process and search for special auxiliaries in combinations
- process and search notation built through parallel division

c) process and search ranges of UDC notation - thereby making it possible for a computer to generate each number within a specified range in order to enable the retrieval of the concepts covered by a range of numbers. Ranges in the UDC occur in main numbers, special auxiliaries and main auxiliaries on any level of division. Buxton (1990: 199) specifies the following situations that occur in ranges and for which the following algorithms need to be performed by programs:

- aaa/xxx - generate each number between aaa/xxx unless final digits in aaa or xxx are 0 (zero)
- aaa.bb/.cc - generate each number between aaa.bb and aaa.cc
- aaa.b/.cc - generate each number between aaa.b and aaa.c and generate aaa.cc

3) full access and control of all the occurrences of a notation in different combinations in order to change globally or manage them centrally (export, link/trace, print, add notes, and display).
4) control over notation types (single, combined), tables and number syntax for the automation of filing or number building (citation order, filing order)

Based on these four groups of requirements, it is obvious that control and access over each element in a pre-combined number is central for UDC number processing. It is not just necessary, however, to be able to **search** for synthesised numbers, which can be achieved through a programmatic approach of UDC number decomposition in the process of IR, as explained in Chapter Four. Other important functions are the **management** of display, sorting/filing and global changes of data, and for this purpose it is necessary that the synthesised UDC headings are properly labelled for database processing. This then enables the implementation of a larger number of important operations:

- searching of composite elements
- changing the notation display, e.g. different colours for main auxiliaries and main numbers, or different symbols (in case IR systems have problems handling the UDC default)
- global changing in the process of reclassification
- sorting notation independently of symbols and controlling the sort and display process, e.g. re-grouping and displaying of UDC numbers by a certain auxiliary number irrespective of the general rules of citation order

In the past there have been many suggestions that UDC notation should be more formalised and simplified (see Chapter One, section 1.2). Recent objections to the inconsistent use of notation in the UDC were put forward by Wimmer (2000) in relation to the apostrophe. Strachan suggested that automatic meaningless points should be abolished while some notational symbols such as .0 and the asterisk should be replaced with other symbols to ease their processing (Strachan, 2000). Some changes can be made, and indeed have already been made, in the process of revision such as the removal of the *final digit* in 2000.[72] The right solution is not, however, to change the UDC notation or its display but rather in dissociating the way UDC numbers are displayed from the way they are processed. In order to search parts of pre-combined UDC numbers, Buxton (1990: 214-215) has suggested three possible approaches:

- to split synthesised numbers using spaces
- to replace UDC symbols with alphabetical codes whereby each letter would act as an facet indicator
- to introduce subfields and encode composite UDC number as a series of subfields [i.e. use tags]

Only the last of these three recommendations would fully meet the

[72] The change is introduced in *Extensions & Corrections to the UDC* 22, 2000, page 79.

requirements for the full processing and management of UDC synthesised numbers, as was also pointed out by Buxton. If tags for encoding each part of a pre-composed UDC number are used, two further technical solutions should be considered. This is illustrated in example *73.033-032.5-053.2(43) Mediaeval German sculpture in stone representing a child* which consists of the following elements:

73Sculpture
.033Mediaeval[facet indicator .0 meaning special auxiliary)
-032.5Stone[facet indictor -03 meaning common auxiliary of material]
-053.2Child[facet indictor -05 meaning common auxiliary of persons]
(43)Germany[facet indicator (1/9) meaning common auxiliary of place]

The solutions for encoding include:

- code in a way that facet indicators are replaced by tags that represent their meaning and all punctuation is removed so the number *73.033-032.5-053.2(43)* is coded as:

  <TAG1>73<TAG2>33<TAG3>25<TAG4>32<TAG5>43
- code in a way that tags representing the meaning are put in front of each element but the number *73.033-032.5-053.2(43)* itself is represented in its original form separated by tags:

  <TAG1>73<TAG2>.033<TAG3>-032.5<TAG4>-053.2<TAG5> (43)

In the first case, UDC notations are altered and stripped of any symbols that may pose a problem for data handling. Special instructions are then written so that in any display, editing or export of the data these tags are replaced by the original UDC facet symbols. In the process of searching, when users are using symbols, every search needs to be handled by an additional program that will remove this punctuation and replace it with adequate tagging in order for database interrogation to be performed. One advantage of this is claimed to be the ease of data administration, which does not contain symbols that interfere with database interrogation. The disadvantage of this approach is the amount of additional programming involved. Also, as the original UDC data have been changed, their preservation and future restoration may be jeopardized by dependency on an external coding system.

The second approach is considered to have disadvantages in data administration as, for instance, SQL statements have to escape special characters in searching or sorting. Also, there is redundancy as facet indicators such as .0, -03, -05, or () repeat what is already stated by the tag. In the process of searching, however, there would be no need to translate the search statement for database querying and also, as the data are kept in the original form, no program intervention is needed in the display, print or export or data.

If either of these approaches is to be adopted, the best method would be to store the UDC number representations in separate fields: one retaining the original text string and the other holding the structured/tagged notation:

```
<Field 1>    73.033-032.5-053.2(43)
<Field 2>    <TAG1>73<TAG2>.033<TAG3>-032.5<TAG4>-053.2<TAG5> (43)
```

In terms of database management, the first field would be used for complex notation display, string search purposes and for the preservation of original data. The second would support automated functions (searching, displaying elements of pre-combined numbers sorting, global changes etc). Since the UDC MRF does not offer any solution for pre-combined number coding (i.e. no approach to processing UDC data is presented in the original data source) the solution is still left to individual implementers.

6.2.2 Hierarchical and associative relationships

Semantic or paradigmatic relationships (hierarchical and associative) are the classification's greatest assets with respect to IR. Making these relationships functional would enable linear and non-linear browsing and search expansion, which may be used for improving recall and building a search terms dictionary/index.

6.2.2.1 Hierarchical relationships

The UDC's hierarchical order is to a large extent made explicit through its notation. Occasionally, however, the hierarchy may be completely misinterpreted through notation. This is the reason why an automated system cannot rely fully on notation to find broader or narrower classes; hence the need for data on hierarchical relationships to be made explicit for machine processing. Effectively this means that there is a need for a separate data element that can hold tracing information for each UDC number's broader class. In automation, special attention should be paid to the cases where the following 'faults' appear in the hierarchy, as they need to be corrected manually:

A 'Hidden' hierarchy (i.e. not expressed in the notation**)** is the type of notation in which subordinated classes are shown as coordinated. This often happens when the class spans a centesimal level (more than ten subdivisions are required) but it also happens when the rule of a 'telescoped hierarchy' or 'telescoped array'[73] is adopted in order to shorten the notation (Ranganathan, 1962). The remedy for this would be to change the broader class notation to span numbers as is described below in 'span number hierarchy', otherwise subclasses will remain coordinated instead of subordinated. It is not possible to detect these situations in UDC automatically as they are not marked in any way and substantial effort has to be made in order to rectify hidden hierarchies or mark for special treatment in hierarchy management. The following example shows class (73) United States before it was revised in 2000:

(73)	United States of America
(74)	North-eastern States
(75)	South-eastern States
(76)	South-central States

Ranganathan (1962: 278) gives the following as an example of good use of a telescoped array in shortening the notation, which he allows to be an exception to his Canon of Hierarchy:

(1) Place and space in general
[beginning of the telescoped array 2 to array 1)
(3) Ancient world
(4) Europe
(5) Asia
(6) Africa
(7) North and Central America
(8) South America
(9) South Pacific and Australia
[end of array]

A 'range number' hierarchy is the use of a range of notations to express broader classes and is a regular feature of the UDC for reasons mentioned under 'hidden hierarchies'. Subclasses, in this case, cannot be automatically linked to their broader class. An illustrative example of a hierarchy, which needs to be 'translated'

[73] Definition of a telescoped array: "*Array of classes in a schedule of classification, made of co-ordinate and subordinate isolates, as viewed from the Idea Plane, but whose class numbers appear to be co-ordinate, as viewed from Notation Plane*" (Ranganathan, 1962: 278).

for machine handling, is class *59 Zoology*:

59	Zoology
591	General zoology
592/599	Systematic zoology
592	Invertebrata
593	[empty]
593.1	Protozoa
594	Mollusca. Shellfish etc.
595	Articulata
596/599	Chordata
596	[empty]
596.2	Urochordata (Tunicata)
597/599	Vertebrata
597.2/.5	Pisces: Fishes. Ichthyology
597.6	Amphibia. Amphibians in general
597.7	Gymnophiona (Apoda). Caecilians
597.8	Anura (Salientia). Tailless amphibians

In the MRF 2003 there are over 900 records containing combined notation with a 'range'.[74]

A 'level jump' hierarchy (i.e. missing levels) appears as a fault in the notation, or as a result of cancelling or superfluous concepts or when a class number is synthesised from a very specific common/special auxiliary and broader levels are not introduced. In this case, several hierarchical levels are missing and need to be added, even if there are no documents classed at these levels, in order to help logical navigation. This can be seen as a fault in assigning notation (the first example) or it can occur when creating synthesised numbers (the second example):

316.776.4Metacommunication
316.8[empty]
316.81[empty]
316.811Marriage and family
316.811.1Nuptiality. Marriage. Union. Cohabitation
316.811.11Types of marriage, union, cohabitation
316.811.111Marriage by number of spouses
316.811.112Marriage between partners from different cultural backgrounds
316.811.115Same-sex marriage. Gay marriage

2-282.7 Service Books [in UDC MRF 2003]

271Eastern Church [in UDC MRF 2003]
271.2Orthodox Church [in UDC MRF 2003]
271.2-282.7Service books - Orthodox Church [missing]
271.2-282.7-2Service books - Evidence of religion - Orthodox Church [missing]
271.2-282.7-24Service books - Specific texts - Orthodox Church [missing]
271.2-282.7-247The Gospel Book [in UDC MRF 2003]

[74] When UDC is used for a real collection, some of these range combinations will be used and many new synthesised numbers that do not exist in the MRF will be added in the process of classifying.

These 'holes' in the hierarchy pose problems when implementing a classification for browsing, and are an obstacle when there is a need to broaden the search by truncation as this will produce zero results. The 'holes' in the UDC itself tend to be corrected when discovered while those in synthesised numbers usually occur in the implementation of the classification in a library and tend to be systematically filled in by cataloguers. The problem here is that these cannot be automatically discovered in the MRF but need the help of an additional tool.

A 'False' hierarchy occurs as a mistake or as a result of revision when other solutions cannot be found. Mistakes are likely to be corrected eventually by the revision of the UDC. The hierarchical irregularities that are likely to stay, however, are those caused by a mixture of place numbers for old and new geographical divisions. In the following example *Galicia*, which is the name for the former territories that are now shared between southern *Poland* and western *Ukraine* appears in this notation to be part of the former *Czechoslovakia*. *Bukowina*, which was the southeast part of *Galicia*, does not appear to be part of *Galicia* but rather a part of *Czechoslovakia* while, in effect, at present this is the land between *Romania* and the *Ukraine*.

(437)Czechoslovakia (1918-1992)
(437.3)Czech Republic
(437.4)Galicia (Galizien) to 1919
(437.5)Bukowina to 1919
(437.6)Slovakia. Slovak Republic. Slovenská Republika
(437.7)Zakarpatska Ukrajina (Transcarpathian Ukraine/Carpatho Russia/Ruthenia), 1918-1938

Because of the problem in browsing or search expansion that may be caused by the four situations described above, it is obvious that notation is not a reliable indicator of hierarchy and that machine-processable information has to be added to UDC data in order for the hierarchy to be managed automatically (cf. Loth, 1996). The first step in the coding of hierarchies for database management would be to code the hierarchy level, where this can be achieved automatically and, to code manually all situations where the notation does not express a hierarchy (i.e. the four situations above). In relational databases holding UDC data, a table holding hierarchy codes may be devised.

Table 6.1 shows how a notation that has six levels from *316 Sociology* i.e. *316.811* (shown in the example of 'level jump' hierarchy on the previous page) is actually coded as four level notation 1.1.1.1, which will result in the appropriate display (indentation) and search expansion (broadening) in an online system. Also, each four level notation is assigned a filing sequence number within the whole database that allows full control of the hierarchy at the whole database level. Such coding enables collapsing/expanding functions on the interface, or the management and export of the UDC by levels of division.

Table 6.1: Example of database, notation independent, hierarchy coding

UDC NUMBER	UDC RECORD ID	HIERARCHY LEVEL AND FILING ORDER
3	016994	1
31	017253	1.1
316	017332	1.1.1
316.811	017681	1.1.1.1
316.811.1	017682	1.1.1.1.1
316.811.11	017683	1.1.1.1.1.1
316.811.111	017684	1.1.1.1.1.1.1
316.811.112	017685	1.1.1.1.1.1.2
316.811.115	017686	1.1.1.1.1.1.3
316.812	017687	1.1.1.2

Because of the inevitability of 'faults' in notational hierarchy, the basic requirement in UDC management and maintenance is to rectify this by recording, permanently keeping and maintaining data on a broader class for each UDC number as an additional data element. This will help reduce the effort required in correcting the hierarchy in any future implementation of the UDC and will help in the automatic development of coding as shown in Table 6.1.

Hierarchical relationships that occur in the classification are expected to be *generic* or *partitive*, but in spite of revision efforts to enforce this, they may also exist as a less desirable *member/class* and even as *class/instance* relationships. Svenonius (1997) pointed out that the mixture and impurity of these relationships may even be an advantage, as they will reveal unexpected but pertinent semantic relationships based on literary warrant that may be useful in IR. For the full exploitation of KOS in automation and especially in expert systems (automatic indexing) it would, however, be desirable to have the type of hierarchical relationship properly marked, so they can be better exploited in an m2m environment. The enrichment of UDC data with a hierarchy type (broader/generic, broader/partitive or otherwise), is not a basic requirement and it is a matter for classification developers to consider if they wish to add more value to UDC data. When the UDC MRF contains appropriate data on hierarchical relationships, UDC users will save time and the duplication of effort in implementing the scheme.

6.2.2.2 Associative relationships

'Lateral' relationships exist in the UDC as '*see also*' and '*see*' references. While the former are mostly encoded in the MRF in the references field (field 125 of the MRF database), the latter can be found in the application note or in the 'parallel division' instructions. *See also* references are an important way of linking 'distributed relatives' in a classification structure. As opposed to a vertical i.e. a tree structure of hierarchical relationships, the syndetic structure of *see also* references represents complementary semantic linking in the form of a lattice. If properly automated (hyperlinked), references are very important in IR as they provide points of semantic expansion and allow for lateral browsing. As in other classifications, these relationships represent a 'mixed bag' of all other semantic relationships that are not hierarchical.[75] In the context of advanced IR models and the Semantic Web, the value of a controlled vocabulary would be more significant if it had associative relationships defined, encoded and controlled. The problem is that in many cases, as pointed out by Svenonius (1982), it is not possible to verbalise or to categorise some of the associative relationships as they are not definitional in nature.[76]

Although it is not easy or straightforward, it is still possible to differentiate between some associative relationships and many were advocated for their usefulness (Ranganathan, 1967; Farradane, 1961, Perreault, 1969). Modern thesauri, such as e.g. AAT attempt to record all parts and all uses of listed concepts and they attempt to define the rules of recording associative relationships that are relevant for information retrieval (Molholt, 2001). Literature on thesaurus construction also

[75] There has been no strict policy with regards to the assignment of the *see also* references in the revision of the UDC although it was argued that the value of a syndetic structure is such that UDC should significantly increase this kind of class linking (Dahlberg, 1971).
[76] "*Many of the related-term relationships in these tools represent empirical generalisations; that is, by virtue of a sufficient number of past occurrences, two words come to be regarded as associated, e.g. literary warrant may justify associating 'frustration' and 'achievement*'" (Svenonius, 1982 :131).

suggests the categorisation of associative relationships (Aitchison, Gilchrist & Bawden, 2000: 62-63).

The structure of a thesaurus is clearly not totally comparable with classification. Some of the relationships that need to be specified in a thesaurus are already fixed in the classification's structure. It should also be noted that not every field of knowledge requires the same number of associative relationships. For instance, it has been acknowledged that it is difficult to establish rules for associative relationships in the humanities and social sciences because of their very pervasive nature. Differentiating between associative relationships and categorising some of them as a general guide would, however, help in validating and introducing useful semantic linking if desired.

On many occasions in the past it has been pointed out that UDC should enrich its syndetic structure and add more associative linking (Lloyd, 1972; Dahlberg; 1971; Task force for UDC system development, 1990). That implies that it was recognized that there are situations in the UDC where associative relationships, although desirable, were not established. For instance, classes *632.9 Control of plant diseases* and *632.95 Plant protection preparations. Pesticides* (or any other sub-classes) are not associated with the class *543.393 Pesticides*. Other examples are classes 364.632 *Physical abuse*. *Violence, -058.68 Victims of ill-treatment* and *343.988 Victimology* are not associated through see also references.

The enrichment of the UDC with *see also* relationships has become more feasible, and significantly easier, since 1993 after the creation of the UDC database. The analysis of the UDC MRF data in Chapter Eight, however, shows that half of the references in MRF 2003 are not cross referenced.

The basic requirements for processing associative relationships could be summarised as follows: *(a)* to encode *see also* references in a consistent manner *(b)* to have **cross referencing** functional and validated *(c)* to have coded 'see' references in annotations and scope notes. Once these parts of UDC data are properly coded, maintained, validated and controlled within the file UDC MRF, the user may find the implementation of see also references in a IR system much easier and cheaper.

6.2.3 Linking notations and verbal expression

The use of classification in browsing and in searching creates the need for two different, co-existent 'translations' of UDC notation: *(a)* UDC caption and *(b)*

UDC search term index (subject-alphabetical index). In relation to the caption, the following needs to be provided for any database supporting UDC display and browsing:

- provision a caption for pre-combined classmarks
- caption adjustment and preparation for its use in IR
  - to extend the caption to become less dependent on hierarchical context and to become more meaningful, complete and self-contained
  - to summarise 'over-explanatory' captions i.e. to remove excessive wording
  - to remove unnecessary abbreviations, comments or explanations from caption
  - to adjust the caption to include terms from subsequent subdivisions if these are discarded

In relation to terms that may be added for the purpose of searching to:

- ensure the term inheritance from single classes to composed classes
- ensure cross linking of all terms describing the class or relating to the class
- enrich the terminology with synonyms and quasi-synonyms
- link distributed relatives through verbal expressions
- contextualise homonyms in term index display
- prioritise the sequence of terms assigned to each class

Searching words obviously includes several steps or actions that may be supported by a vocabulary management system: linking of terms, editing and correcting the existing vocabulary, adding new terms, and attributing a relevance ranking to all terms assigned to a single class. This is the reason why it is important to have access, control and appropriate independent management of both kinds of terminology attached to the UDC. As this was the case with other issues in processing UDC data and its implementation, the most desirable and appropriate solution would be for the UDC MRF (which is the main source of UDC data) to contain index terms related to each notation as well as for the captions to be completely consistent in style and form throughout the UDC.

6.3 Classification management and authority control

Authority control is an expression used by the library community to denote a procedure that ensures that every new cataloguing description will contain the same approved form of name, term or classification number. Authority control can be based on a file in an isolated system or can be created, exchanged or shared among networked or federated systems to reduce resource waste as well as to enable searching across systems.

Authority control can be described as a both a procedure and a function for establishing access points and linking all the forms of a term within a single authority record connected to a real document collection. Initially, authority files were created to hold a single term, name or classification number chosen to serve as an access point (heading). Gradually, with the development of library systems, the function of an authority file was extended to hold not only the approved headings but all variants of possible access points leading to the main heading as well as any other data necessary for system management and information retrieval. Markey Drabenstott (1996a) made an important point about the implication of authority control being both a language for describing user preferences and a language for describing collection content.

Authority control that holds all relevant access points (for users and for librarians) also serves as an effective **access control** and has great importance in information exchange between networked systems (Barnhart, 1996). The economy of library systems: the need for control, validation and re-usability of data made subject authority files an important tool in bibliographic systems fulfilling three functions: *(a)* bibliographic database maintenance, *(b)* subject analysis and indexing and *(c)* subject access (Mandel, 1995). Tillett (1995) depicted the rôle of authority control in the 21st century as paramount in improving access through linking large bibliographic databases which will work by improving system architecture and navigational methods using complex linking between bibliographic descriptions and authority control files. In such a model, as emphasised by Barnhart (1996), an authority file enables greater automation of the cataloguing/classification process as well as greater precision and recall in information retrieval as it becomes an access control tool. Authority files, which hold contextual information, are fundamental for any kind of semantic interoperability in a networked environment and libraries are part of such an information space (Tillett, 2000; Cordeiro, 2003).

In the case of classification, an authority record will contain a classification number as well as all the information about it: its hierarchy, description, search terms and the relationships to other classmarks, the rules for its use and, in this case, an authority file serves as an authority control and classification tool. Every new notation that needs to be introduced is first entered and described in the authority file and then linked to the document; all other document descriptions that ought to share the same notation are simply linked to the previously established notation authority record (illustrated with Figure 6.2 in section 6.3.3). It is clear, therefore, how important such a tool can be in supporting both subject indexing and information retrieval. Apart from its importance for the local management of an information system, a classification authority file is even more important in information exchange and discovery as it can be shared between different library systems (Mandel, 1995).

6.3.1 UDC schedules vs. UDC authority file

A classification authority file may contain all the classmarks already present in the original classification schedules but, more importantly, if compared to classification schedules kept in an electronic form, such as the UDC MRF, it holds additional data necessary to manage and use classification in information retrieval in a real bibliographic system (Table 6.2).

Table 6.2: Data elements in classification schedules and authority file

UDC DATA	UDC MRF	AUTHORITY FILE
simple UDC numbers	•	•
authorised selection of pre-synthesised numbers	•	•
caption	•	•
relationships (hierarchical and associative)	•	•
instructions and rules for combination	•	•
different notes (scope note, editorial note)	•	•
pre-synthesised UDC headings	(•)	•
history notes, indexing policy notes, reclassification notes and indexing instructions		•
locally assigned search terms in one or more languages		•
mappings to other indexing languages such as subject heading systems or other classifications		•
locally assigned see also references		•
links to actual bibliographic references of real documents and examples		•
tracing of hierarchical relationships		•

With the automation of classification schedules (DDC, UDC and then LCC) for the purpose of their maintenance, distribution and publishing, the differences in the rôle of *automated classification schedules* and a *classification authority file* have become more obvious and more relevant. Schedules contain a standard choice of concepts and their combinations. Authority files, on the other hand, can contain an unlimited and unpredictable combination of concepts as they occur in documents in a real and growing collection.

The differences between electronic classification schedules and authority files increase with the system's affinity for number building. Enumerative classifications offer a limited possibility to synthesise classmarks and an authority file will offer only a relatively small percentage of additional pre-coordinated classmark strings that are not listed in the schedules. For UDC, however, number building is the underlying principle of the whole system and an authority file containing a list of pre-composed classmarks will represent a significantly different and unique 'universe of knowledge' that pertains to a specific collection. UDC schedules allow for great specificity and exhaustivity in indexing through number building. This makes the authority file an important tool in classification as it will facilitate the re-use of previously built numbers, control subject aggregation and prevent document scattering.

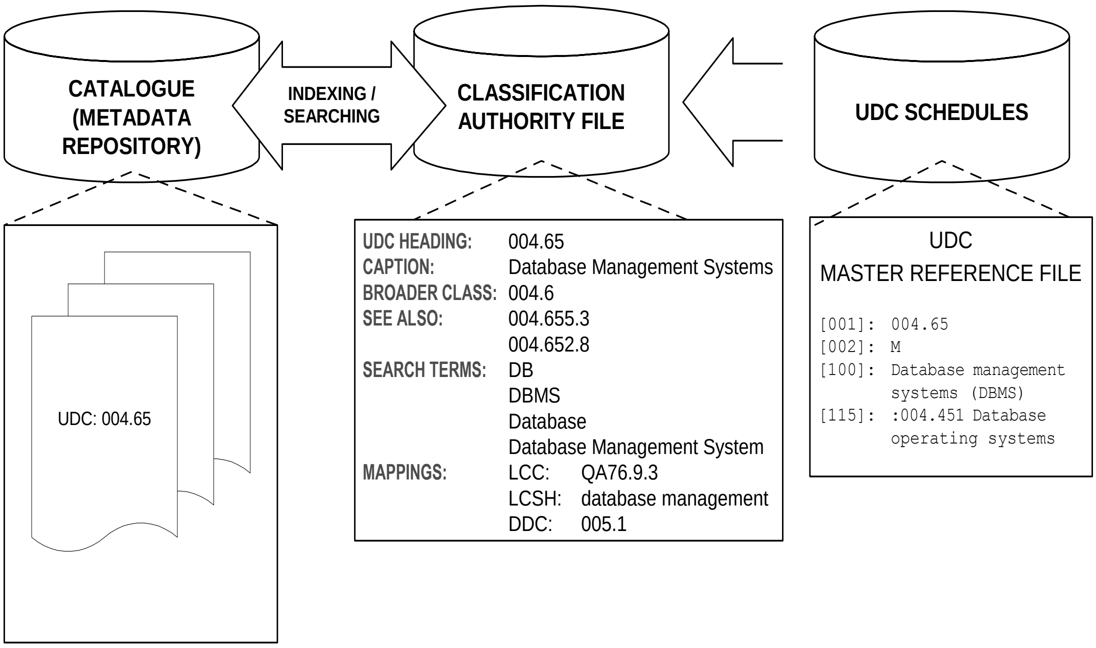


Figure 6.1: Relationship between UDC, authority file and bibliographic databases

The added value of an authority file, if compared to classification schedules, is that it contains pre-constructed strings of classification headings based on literary warrant as well as a choice of search terms. Large libraries with mature and carefully maintained UDC authority files possess a tool that may have value for many libraries, which are at the beginning of the process (see Figure 6.1). Because of the work involved in the creation and maintenance of a classification authority file, it is helpful if the classification schedules offer high quality data that can be easily incorporated into the authority file or be used for its update.

The exchange or sharing of authority data is closely dependent on the availability of international standards that would normalize the form, format and content of authority records. Since the early 1980s, libraries have been putting pressure on library system vendors to provide basic functionality that will enable cataloguers to create and manage authority files (Tillett, 1995; Taylor A. G., 1999). The vendors' response was closely dependent on library standards, that is, the only data formats that were published as standards were supported.[77] For one reason or another, the process of developing standards and formats for authority control as well as their full implementation in library systems was unexpectedly slow (Taylor A. G., 1999a). Every classification authority file is more successful and easy to implement and maintain if the original data of the classification schedules come with a data structure that is compatible with the standardized authority file format. In the bibliographic domain this compatibility can be resolved only through the adherence of UDC MRF to MARC formats.

6.3.2 Classification authority file formats

While it could be said that the authority control of names is widely

---

[77] Historically, authority control development in the library world is usually linked to the Paris Principles in 1961 (A.G. Taylor, 1999). In the four decades following, a number of relevant standards for authority control were created under the umbrella of the IFLA Programme for Universal Bibliographic Control (UBC) or by the Library of Congress. A relevant standard was Guidelines for Authority and Reference Entries (GARE) in 1984. This was followed by formats for machine-readable authority data: USMARC format for authority data in 1981 and *UNIMARC/Authorities* in 1989-1991 (Cordeiro, 2003). The Functional Requirements and Numbering of Authority Records (FRANAR) that was intended to address the question of a unique identifier for authority records ISADN (International Standard Authority Data Number) was released in 2003.

established and name authority files have been successfully created, shared and exchanged between systems, subject authority control is fairly new and, as described by Mandel, represents 'uncharted territory' with unknown functionality, costs and benefits (Mandel, 1995). The lack of MARC formats to support the creation and exchange of subject authority records was the greatest obstacle in this respect. The work on international standards to support subject authorities, however, started fairly late: the USMARC format for classification data was published in 1990, GSARE - *Guidelines for Subject Authority and Reference Entries* was issued by International Federation of Library Associations (IFLA) in 1993; and the *UNIMARC Classification Format*, based on USMARC was published in 2000 (in an, as yet, unfinished form).

The reasons behind the slow development in this area were the lack of support for subject searching in library systems and the differences between indexing languages in terms of their nature, structure, complexity and functionality. The creation of a universal classification standard format was certainly impeded by the structural differences and competitiveness between different classification systems, which were simultaneously undergoing different stages of their own automation. Institutions and bodies in a position to influence standards such as the Library of Congress and OCLC were logically concerned with the subject vocabularies they maintained, used and promoted and did not have much incentive to think about generic standards that could serve as vehicles of exchange and use by other systems. This bias was further encouraged by a lack of conceptual data models for each particular classification system and consequently meta-models that would work for more classification systems (Cordeiro & Slavic, 2002).

6.3.3 The issues of synthetic classification

Pre-coordinated indexing languages, and especially classifications, were often rejected for being costly and time-consuming. This was in spite of suggestions that these systems could be used efficiently and to their full potential once cataloguers were supported by adequate computer applications. Semi-automated systems help cataloguers to choose terms quickly, consistently and accurately and to establish their coordination but they also provide more access points and better retrieval techniques (Svenonius, 1992a). Fugmann (1990) also argued that highly expressive syntactical notation ought to be managed by computers.

Properly automated classification, for instance, can establish and preserve vocabulary transparency. Furthermore, the success in indexing using analytico-synthetic classifications may be more dependent on machine-assisted indexing as synthetic classification notation is more complex, formalised and therefore less suitable for manual editing. One of the most important requirements for the management of analytico-synthetic classifications for information retrieval is the management of synthesised notation and any standard, format or tagging scheme used in a classification database should be judged by its capability to encode the structure of synthesised numbers.

In the case of the UDC, it is helpful that classification structural data elements are already, to a great extent, evident from the system's notation and this can assist in designing formats for their management. For instance, numbers from facets of common auxiliaries or special auxiliaries are differentiated through prefix symbols, while single main numbers do not have any prefix and do not contain .0 or any other symbol apart from a mechanical point.

Synthesis of UDC notation may be achieved through the straightforward aggregation of notations from different facets or by combining two notations with a relational symbol. The resulting UDC synthesised notations will contain composite units separated by facet indicators (prefix symbols) and/or relationship symbols which make them easy to recognize by the human eye. Conventions and rules about the meaning of these symbols, and rules for their use, are recorded at the classification system level and they form the body of knowledge necessary to use the system.[78]

The automation of a classification is successful only if the same 'transparency' and 'system rules' are made available for machine processing and this usually means that data representation in the form of notation ought to be further formalised and formatted using more detailed encoding schemes (Slavic & Cordeiro, 2004; Slavic & Cordeiro, 2004a).

---

[78] The conceptual, faceted structure of main fields of knowledge when present in UDC, however, is not made expressive through notation and is not amenable to control. For instance, very often facets of agents, processes, operations or material are denoted as simple main numbers and the beginning/ending of the facet is not marked in any way in the schedules or in the UDC MRF database. This approach means that facet analysis is not made functional in facilitating or controlling concept combinations.

The importance of encoding *facet indicators* is often emphasised as a requirement in the literature on faceted classification automation (Gödert, 1991; Gopinath & Prasad, 1994; Madalli & Prasad, 2002). This point could be further reinforced by the fact that the need for coding components of a classmark for database management was expressed even for enumerative classifications such as DDC.

With the increasing use of automated systems, DDC owners and developers, for instance, became very eager to make use of any synthetic feature in their structure. This approach is so important that the coding of vocabulary facets, whenever these can be made explicit, is advocated by both those who develop this system and those who are implementing it (Wajenberg, 1983; Pollitt & Tinker, 2000; Liu, 1996; Mitchell, 1997; Miksa, 1998).

There is an important parallelism in the management of all pre-coordinated indexing languages, and not only classifications, as the ease of their use and exploitation rests on the way they are stored and managed in databases. Authority files are used to hold pre-combined index terms that can be easily re-used or searched.

When arguing the need for a more structured subject heading authority file in 1992, Markey Drabenstott pointed out that the library community had long expressed this requirement. She pointed out that there has been lacking machine-readable authority data to aid subject heading assignment and validation since the 1980s. She emphasised the limitation of existing authority records which generally contain headings that are not subdivided; this ruled out the possibility to access, manipulate and search for individual components (Markey Drabenstott, 1992). There is an obvious parallelism between analytico-synthetic classifications and subject headings systems that require the same machine assistance in formulating and validating pre-composed indexes.

Figure 6.2 illustrates how a classification authority file that holds parts of the complex UDC notation coded individually helps control access and searching for any component of a number, regardless of its position such as (091) at the end or (460) in the middle or at the end of the complex number string.

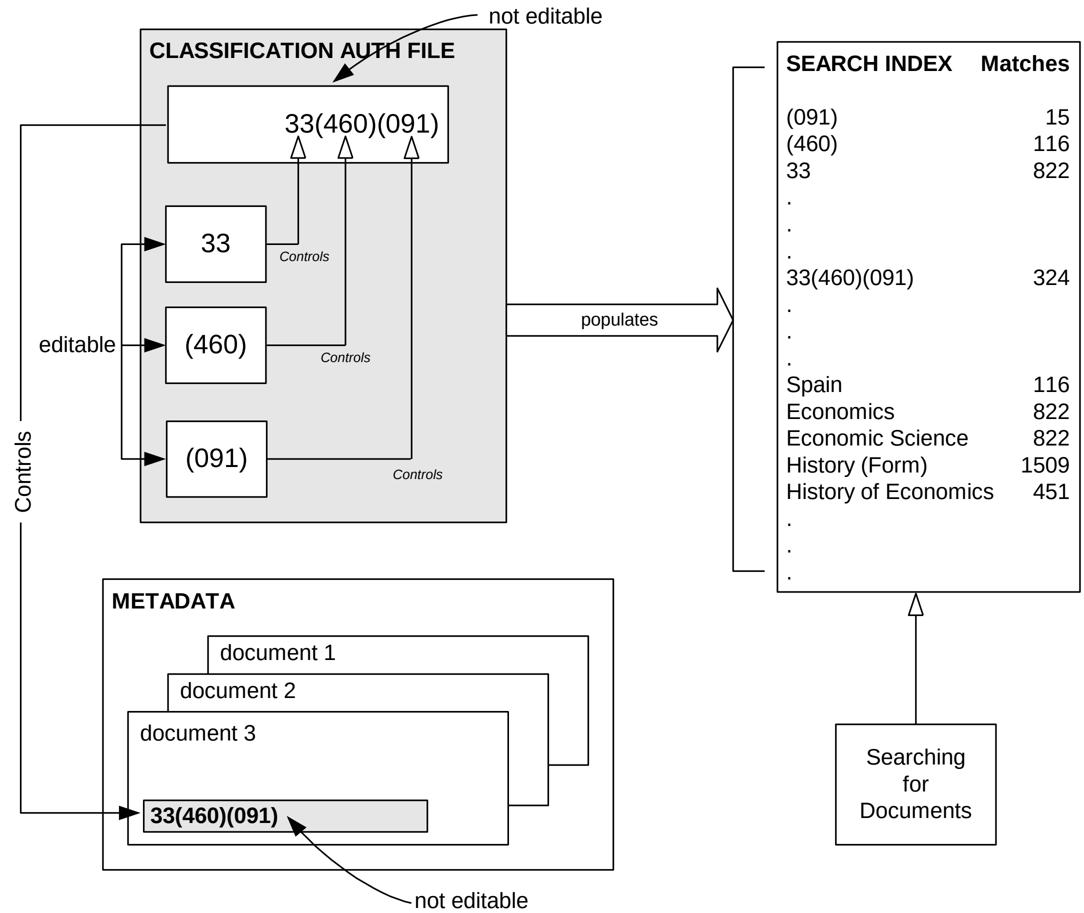


Figure 6.2: Searching for the elements of synthesised numbers

This multidirectional access to every semantically relevant unit is very important for the use of a classification authority file in the searching, display/presentation and sorting of notations, as explained earlier in this chapter and also in Chapter Four.

Figure 6.3 illustrates yet another important virtue of the separate coding of each element in complex numbers, which is very useful in authority file maintenance and reclassification. The same approach to coding allows for global changes, which means that a single editing of a record in an authority file containing a simple number will automatically change all the occurrences of that number in different combinations in the authority file and in the metadata.

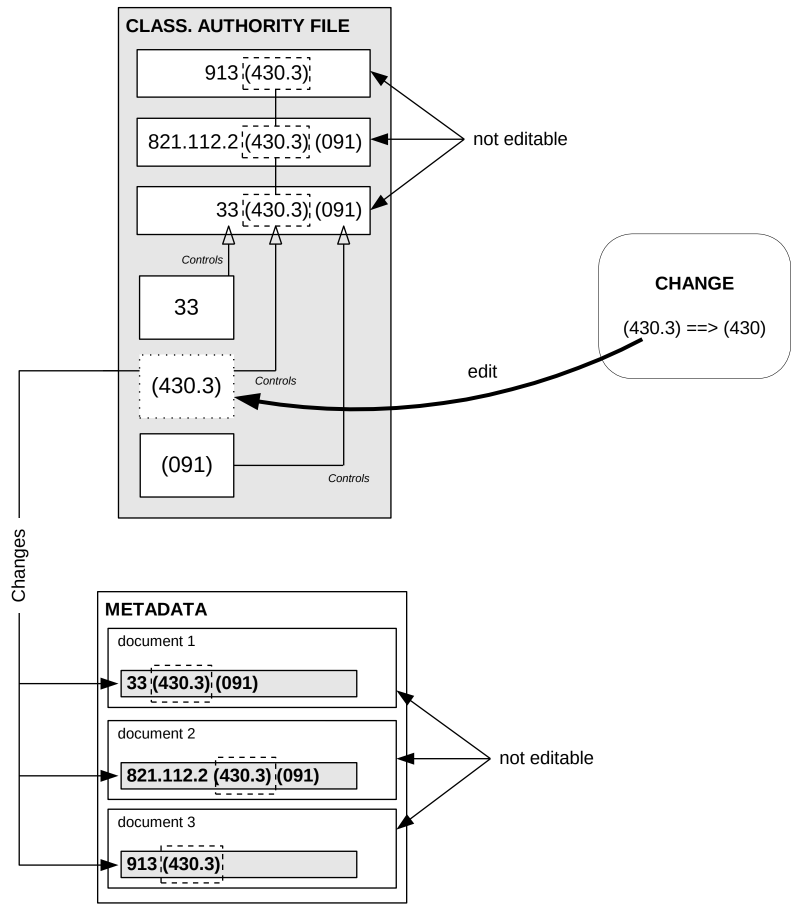


Figure 6.3: Global changes

These two points demonstrate the basic functions of a classification authority tool in the management of synthesised notation. The required functionality of an authority tool can be achieved through the proper analysis of classification data and separation of data content/function from its representation. An appropriately detailed tagging scheme is fundamental as it will permit data formatting that will preserve the semantic and functional value of all original notational components.

6.3.4 Standard formats for classification authority control

It was the maturity of library systems, combined with improving conditions for interaction, which allowed the management of classification to find a place on the standards creation agenda. With the UNIMARC format an attempt was made in the second half of the 1990s to find common ground for the encoding of different classification systems in order to have them maintained, used and exchanged within a community of library users. In this context classification data management and authority control became a single issue with two purposes: the first is holding and managing a classification in a local system for the purpose of cataloguing and information retrieval, the second is that of information exchange. In the case of the UDC the first purpose could be fulfilled if the format were compatible with the requirements for managing and searching UDC numbers while the second purpose is fulfilled if the data format complies with existing community standards e.g. MARC.

The implementation of a **classification** in the library domain requires that a number of choices and levels of use need to be taken into account with regard to database format standards. When applied in a library system, classification notations may appear as MARC format, encoded in two related ways: *(a)* classification notations in the bibliographic records and *(b)* classification notations in the classification authority file (in which each classification notation is fully described and made available for linking to a bibliographic record).

IR retrieval in a library catalogue can be based either on indexes produced from the authority file or those produced from a classification field in the bibliographic record. Another issue is that of the use of the classification in: *(a)* systematic shelf arrangement, *(b)* detailed content indexing for IR[79] and *(c)* the use of two parallel but different ways of classification to meet both functions. When this is the case a classification authority file ought to provide for two parallel functions: shelf arrangement or a shelflist catalogue, and subject arrangement or what is traditionally called a systematic or classified catalogue.

In addition, there are different implementation practices in which a classification authority file has to meet the needs of: *(a)* libraries that use classification simply as a 'mark and park' device with no associated IR function and no number encoding and *(b)* libraries that exploit classification for browsing and post-coordinated searching that required notation to be fully searchable and encoded.

Libraries are the most natural implementation environment for library

---

[79] Classification numbers used for shelf arrangement may be **shorter** or **different** from the numbers used for subject indexing of the very same document. For instance a book containing selected works of Hawaiian poetry, fiction and drama could be marked on a shelf as miscellanea and anthology of e.g. *Hawaiian literature 821.622.821.2-8(082.2)* while classification numbers used to index the content may be described with three subsequent classification numbers *821.622.821.2-1(082.2) Hawaiian poetry (selected works)*; *821.622.821.2-2(082.2) Hawaiian (selected works)*; *821. 622.821.2-3(082.2) Hawaiian prose (selected works)*.

classification, and one would expect them to be the first to tackle the technical issues of classification automation. The MARC bibliographic formats have been, however, processing classification in the same way as would apply to card catalogues. Needless to say, this is not suitable for supporting the parallelisms mentioned above in the needs of a library. In the designated field of a bibliographic record, a classification notation is entered as simple text and each MARC format provides a single coded field to carry such data. Hence, the only way of searching the classification is through an inverted file containing classification notation strings. There are two problems with this: firstly such an inverted file contains only notations and no other access points to the classification (e.g. caption, subject index). The second problem is the fact that the pre-composed notations themselves were searchable only as a whole string or a string truncated from the right.

Wajenberg (1983) was the first to note a discrepancy between the composite structure and dense classification notation semantics and their primitive processing in library systems that cause data loss and inadequacy within classification searching. In 1983, he recommended extended MARC codes for DDC so as to exploit shelfmarks for improving subject searching of bibliographic data.

His idea was that all the data necessary to search DDC would be stored in a bibliographic record in the form of a series of subfields. The required subfields would encode hierarchical relationships and synthetic notations which, in DDC are not discernible from the numbers themselves. He suggested more than ten subfields to encode the total number of combinations. The greatest criticism of his approach was that the full encoding of a classmark needed to be keyed in for each bibliographic record in the process of cataloguing. Wajenberg's argument was that MARC did not hesitate to introduce elaborate patterns elsewhere in the bibliographic records and there was no valid reason not to do so with classification data.

Wajenberg's suggestion was ahead of its time and was independent of the model of an authority control as an externally managed file that would naturally enable re-use of existing data and would not require editing of notation in the process of cataloguing. Cochrane & Markey (1985) correctly anticipated that an appropriate solution would be in classification authority control. Still, a whole decade later, when looking for justification for his DDC automatic decomposition project, Liu (1996: 2) expressed concern about cataloguers' efforts: *"The approach Wajenberg suggested would work, but it would require a great deal of time and effort for catalogers to* *analyze and encode components of synthesised numbers."*

At the time of Wajenberg's proposal, no library classification systems were held, maintained or distributed in electronic form and the next big step was made by Cochrane & Markey (1985) who started their research on classification use online. They were the first to provide an exhaustive list of classification data elements needed for automation (DDC) that went far beyond the number of elements suggested by Wajenberg. This represented an authority control framework independent of the bibliographic description that was capable of serving classification maintenance, distribution and its subsequent use in IR.

With respect to content, classification formats should be able to identify and declare the classification system in use, its version (edition) and language. They should serve as a carrier of classification data so as to encode all the structural and semantic data relevant for indexing and provide access points for information retrieval. Standard formats aim to be as widely applicable as possible and therefore need to be generic on the level of data elements, data types and data structure with respect to the content they carry. This means that classification file formats may achieve a great level of universality by being unbiased with respect the classification system to which they serve as a carrier.

6.3.5 MARC classification formats

A MARC group of formats consists of bibliographic, authority, holdings and community standards. Initially, the only provision for classification in MARC was the classification field within a bibliographic record that limited the use of the classification, and the only way forward was to move towards an authority control tool that would enable classification re-use, validation and control and management.

The MARC format, if used to hold these data, provides additional advantages by being a carrier of bibliographic data in a well-connected community of users. As library systems are designed to process and interpret MARC formats the integration of classification data in this structure was deemed very important for classification exchange and shareability (Guenther, 1996). When Cochrane & Markey (1985) first proposed data elements for machine-readable classification data for DDC they were thinking of the benefits for editors and publishers of the classification as well as library staff and patrons.

General formats for authority data created in the 1980s and the 1990s (the

USMARC Format for Authorities and UNIMARC/Authorities) were not, however, suitable for accommodating classifications. Therefore, the USMARC Format for Classification data was created in 1991 as an independent format to meet the needs of classification authority data and had little overlap with the general format for USMARC authorities. The UNIMARC classification format was developed a decade later, and it was supposed to fit the classification format within the standard of UNIMARC authorities. Any machine-readable classification data could potentially serve as authority data but having a classification in MARC format was a big step towards interoperability which is why Markey Drabenstott called it a 'vehicle' for expressing and sharing classification data (Markey Drabenstott, 1996: 108).[80]

The level of granularity of the coding of data elements is driven by the functionality that the format has to provide and logically each classification element whose manipulation will support searching, maintenance or display is coded as a separate element. The most important function of the format is the successful separation of classification data content and structure from the way data are displayed in print or in an online system.

6.3.5.1 MARC 21 Classification format

USMARC, now renamed to MARC 21 Classification Format[81] was the first standard format created to hold classification data in a full and comprehensive manner. It was stated that the format was created with the intention of being generic and unbiased towards the classification system (USMARC format for classification data, 1997; Guenther, 1992; Guenther, 1994; Guenther, 1996). Its applicability to different classification systems is, however, compromised by the fact that its design was based on DDC and LCC, which are essentially enumerative systems. The format's primary purpose was to provide specifications for converting classification

[80] In addition to classification exchange, these formats can be used for the automation of schedules of classification systems that are not already available in a database format. In 1993 the USMARC format for classification data was, for instance, experimentally used for the conversion of LCC (Guenther, 1996).

[81] Since its creation in 1990, and subsequent publication in 1995, the USMARC format has been regularly updated. The format has been published on the web in a concise form – a quick reference guide (http://www.loc.gov/marc/classification/eccdhome.html) and is being updated once a year since 2000. Changes are indicated by text in red. Examples in this text are made based on the 2003 update.

data into a machine-readable form for communicating and storing classification systems. But as each classification number was stored as a separate USMARC classification record, which contains not only numbers and their captions but also notes associated with their use, this format represented the first standardized authority for classification numbers (Guenther, 1994).

The format provides codes and conventions (tags, indicators, subfield codes and coded values) that identify the elements of a classification record to be used within the USMARC standard. Classification data elements are encoded using the general convention of USMARC and apart from the field elements holding the classification, the format provides other data necessary for linking and controlling the use of classification within the system such as: the kind of record **(008/06)**, classification scheme and edition **(084)** and classification notation general information **(X53)** e.g. a tracing field that serves to directly connect two classification numbers.

Since it is based on LCC and DDC, the distinction is made between three types of classification number (identified in **008/07**): a single number, a defined number span (a range of classification numbers where the beginning and ending are split into two fields and the description of the numbers is a series of captions held in a separate table) and a summary number span (a range of numbers that makes one topic/subject which can be described by one caption).

The format is comprehensive, detailed and very powerful in supporting not only a classification caption but also the requirements typical for the use and management of DDC and LCC. The MARC 21 Format outline is shown in Appendix 5, Table 5.1. MARC 21, as any other MARC format, allows hierarchical linking of records through tracing fields. This means that although the field element holding the broader class notation does not exist in the format as such, a notation independent and machine-readable data on the correct notation hierarchy can be provided.

There are two data fields that are central if the format is to be considered for use with a synthetic classification: field **153 Classification number** and field **765 Synthesised Number Components**. Field **153** is the classification heading and its syntax treats any number as a simple main number and does not provide coding for any building components, apart from number range (span). For instance, in the UDC synthesised number *82:32* [The relationship between] *Politics* and *Literature*, the coding would be as follows:

153##$a**82:32** $hLiteratures$hPolitics$jThe relationship between politics and literature

In this case classification number *32* cannot be accessed or searched and the symbol for relationship is not machine-processable at all.

If the notation is a synthesis of two consecutive numbers representing a 'span' of notations in UDC *321.6/.8 Modern forms of government,* this can be coded using subfields **$a** for the beginning of the 'span' and subfield **$c** for the ending of the 'span':

153##$a**321.6**$c**321.8**$jModern forms of government$hUndemocratic forms of government. Dictatorship in general$hDemocracy$hSupranational forms

All other types of combination of the classification notation in this format are, however, treated differently and they can be described in a separate field **765 Synthesised Number Components**, but the subfield coding schema does not provide the possibility of the unique identification for building components or their access and control in the same way as is done in field **153** for a number's span.

The fact that all the synthesis, apart from the range of main numbers, is described as text, in separate fields and not in the heading, reflects the specific structure of DDC and LCC where synthesis is limited and occurs only as the result of specific instructions given to a specific notation. Elements of synthesised notations (apart from when they form a span of two main numbers) in these two classifications do not have a permanent value or permanent data representation and they change their meaning depending on the base number to which they are attached. Hence, the retrieval of the 'number building' elements does not make sense if they are not traced to their source or to their base number, and the whole point of field **765** is solely to point to the element's original table and to what they are attached to, rather than to identify the elements themselves.

The function of field **765**, as is defined in the format, gives "*information about how a synthesised number or a portion of a synthesised number was built. It traces the different components of a synthesised number, showing the different portions of the number and under which number the add instructions are given. If a number was built using two or more instructions, a separate field 765 is given for* *each instruction. The information in this field is primarily intended to serve as a tracing of how classification numbers are synthesised to assist classifiers. It facilitates computer manipulation of synthesised numbers, both for validation that the numbers have been synthesised correctly and for index-building, to allow searching every use of a specific number*" (MARC 21, 2003).

The possibility to look for these kinds of component (that change their meaning and form depending on to what they are attached) in DDC and LCC is seen as function suitable for librarians only. This is why it is coded in a way that is important for classification 'number building' tools rather than for information retrieval. The components that are meant to be searched in information retrieval such as 'span' numbers are fully coded in the main field **153 Classification number**.

Thus, the main problem with the MARC 21 classification format, if it is to be considered for use with synthetic classifications, is its limitation in controlling and identifying elements of classifications and the variety of relationships that simply do not exist in classifications such as DDC and LCC and for which there are no provisions. These include unlimited combination of numbers using relational symbols and unlimited direct syntheses of numbers from different facets.

In both cases it is necessary to be able to manage, access, and retrieve the notation building elements (numbers) themselves, the type of numbers (facets) as well as the type of relationships. Although MARC 21 can support partial coding of a synthesised notation expressing a range (span) - in the classification heading field **153**, this synthetic facility is envisaged for use with main numbers only and presents a problem for the UDC where a range can occur in common auxiliaries, special auxiliaries or main numbers.

It is not unusual to have in the UDC, for instance, a span between two main numbers followed by a span between common auxiliaries *73/75"13/15"(4/5)* meaning *plastic art, graphic art and painting in the 14th, 15th and 16th century in the countries of Europe and Asia* **or** *73/75.033.1/.033.3(4/5)*, or a span between main numbers, special auxiliaries and common auxiliaries - meaning *Early Christian, Byzantine and Islamic - plastic art, graphic art and painting in the countries of Europe and Asia*.

Additionally, two main numbers in the UDC often engage in relationships symbolised by a colon (:) e.g. *75"17/18":27* meaning the influence of *27 Christian religion* (main number) on 18th and 19th century (time auxiliary) *75 Painting* (main number).

The example below shows a UDC number consisting of a main number *94 History* and a common auxiliary of place *(44) France* whose summary meaning is *94(44) History of France*, but should be accessed and managed through its components. From the fact that *94(44)* is entered as a simple string of characters it is evident that it is not possible to search for or access the second element: *(44).*

```
084 8# $audc$cMRF2003
153 ## $a94(44)$hGeography. Biography History. General history $jHistory of France
753 ## $aHistory
753 ## $aFrance
683 0# $iDetails from$aTable 1(d)
```

In conclusion, if UDC were stored in such a format, it would be possible to code properly only simple notation and spans between two main numbers. Field **765** could be suitable for describing notation building in parallel divisions but would provide little help in all other combinations. As a consequence, all synthesised notations (apart from a range of two main numbers) would be managed and processed as a single text string and data that can be used in retrieval are inevitably lost.

6.3.5.2 UNIMARC Classification Format

Following the example of USMARC, in 1994 an initiative supported by two IFLA standing committees[82] was undertaken to create a corresponding UNIMARC classification format as part of the UNIMARC authority standard. Because UNIMARC was dominant in European library systems where classification and especially UDC is used as the predominant indexing language, the idea was to provide a UNIMARC classification format that would support this kind of authority control.

The Joint Working Group on a Classification Format consisting of experts in DDC, UDC classification and UNIMARC was formed to carry out this work. The first step was to commission a summary of requirements for the UNIMARC

[82] *IFLA Section on Classification and Indexing* and *IFLA Section on Information Technology.*

classification format and this was delivered in its final version in 1996. Requirements for a Format for Classification Data were written by E. H. Woods (1996)[83] ahead of the Joint Working Group.

In the introduction, the author suggested that the document would be focusing mainly on the issue of how to extend the existing USMARC classification format to the UNIMARC format and what extensions would be needed to make the USMARC format applicable for the UDC. The specific problems of this undertaking, however, were not analysed in any detail because, in spite of its very technical title, the document was general and introductory in nature.

A statement on the potential uses of the classification format against which any requirements should be considered exemplifies the depth of the analysis characteristic of this document. These were specified on a very superficial level with somewhat mixed objectives and varying degrees of specificity:

- communication of classification changes to users
- maintenance of classification table
- authority control
- enhancement of subject access in online catalogues
- classifier assistance
- access/use of different classification schemes
- allow use of expert technology
- provide a basis for an online shelflist
- provide the ability to produce subsets of the classification scheme
- accommodate multilingual indexes to the classification scheme

The objectives that embrace publishing and authority control functions were assembled and did not differentiate between the specific requirements needed for the management, development and distribution of classifications and those of the management of classification for information retrieval. Most objectives listed here can be subsumed under either authority control or scheme maintenance and publishing. The problem is that the rudimentary approach taken in this document did

[83] E. H. Woods is an American expert for cataloguing and MARC and the author of several books in the area.

not allow for the inclusion of details important for the reality and technicality of specific classification authority control requirements that would undoubtedly lead to issues of the definition of a generic classification format or *meta-classification schema* that were crucial in creating this standard.

For instance, the relationship between the publishing body on the one hand, and implementers of the classification on the other is much weaker with synthetic classifications such as UDC than with classifications such as DDC and LCC. This is why it is easier to provide the same format for publishing and authority control of enumerative classifications than is the case with synthetic classifications.

Users of synthetic classifications do not depend totally on publishers to introduce new concepts as many concepts may be synthesised from existing schedules. Synthesis, therefore, works on the principle of putting together previously defined established concepts with inherited semantics and they continue to be managed and searched for as independent units, no matter what relationships are established between them in the process of indexing. A synthetic classification has therefore to be supported as both a pre- and post-coordinated system.

The document does not focus sufficiently on the characteristics that differentiate classification systems (LCC, DCC, UDC) that are relevant for automation and denote the difference between MARC 21 and the future UNIMARC classification format. The impression given is the assumption that all the necessary functions are already supported by USMARC and that the UNIMARC format will only be a mere adjustment of the tagging schema to fit a general UNIMARC authority format.

Pre-composed UDC numbers include several types of number building: direct building with auxiliaries (common and special), direct building with parallel division and number building using symbols. It is a basic requirement for all three types of combination to be supported in automation and it is not acceptable or justifiable to discuss a format that 'may include' some and may not include others.

Based on the above-mentioned document, one can assume that field **765** of MARC 21 would somehow perform the function of management and retrieval, while in effect a field of this nature is mainly suitable for describing the rules for number-building using parallel division and possibly for special auxiliary combinations. The lack of a precise analysis of the shortcomings of MARC 21 with respect to the representation and management of synthesised numbers was crucial for further work on the UNIMARC format. Had this been provided in due course, consensus on the format would have been much easier to achieve.

Following the document on requirements, a separate working group was formed to put together the format: *the Permanent UNIMARC Committee (PUC) on UNIMARC Classification Format*. The draft of the UNIMARC classification format that was prepared for discussion and circulated in 1998 and 1999 was in effect a copy of MARC 21 (then still USMARC) with minor changes in the structure and tagging schema.

Four members of the six-member group put forward comments in relation to its unsuitability for synthetic classifications and gave specific suggestions as to how this should be addressed on several occasions in 1999, 2000, and 2001. In October 2000, the comments were not taken onboard and the working group was dismissed.

The format was published in its unfinished form. The issue of unsuitability of this format is still pending and IFLA's Permanent UNIMARC Committee expressed some willingness in 2003 to resolve the issue based on suggestions for formatting synthetic classification by members of the former working group.

The outline of the present Concise UNIMARC Classification Format shown in Appendix 5, Table 5.2 (Concise UNIMARC Format..., 2000). When compared to the MARC 21 outline, it is obvious that the format of data is basically the same, the only difference being the tagging scheme used. Two of the most important fields heading **250 Class number** and **665 Synthesised number** have exactly the same syntax as the corresponding fields in MARC 21 **153 Classification number** and **775 Synthesised Number Components** respectively.

Comparison between two formats (the Concise UNIMARC Classification Format, 2000; and the MARC 21 Concise Format for Classification Data, 2003) is given in Appendix 5, Table 5.3, which shows that text in UNIMARC is a straight copy from MARC 21. The fact that the field structure and tagging schema are the same as MARC 21 implies that no effort was made to create a format more suitable for synthetic classifications.

Failing to discriminate data elements, relationships and types of combination that would provide for both enumerative and synthetic classifications, the format simply copied the solution used for enumerative classifications. An example of coding a synthetic number taken from the UNIMARC format:

250 ##$a**78.071-056.45(=411.16)**$hCreative and interpretative occupations$Musicians and their functions$hSpecial gifted$jJews
665 0#$a**78.071**$b**-056.45**$b**(=411.16)**

The example provided in the format, as presented above, makes obvious the fact that synthesised UDC numbers should be entered in heading field **250** as a simple text string and that in field **665** the number appears to be split by subfields. When the number is split, however, the elements in the combination *78.071-056.45(=411.16)* are designated with **$a** (the number where instructions are found) and **$b** (base numbers). This does not provide any information necessary to manage facets and search; for example ethnic grouping or persons in relation to subject. Also this is a typical example of where subfield **$a** of field **665** meaning **Numbers where the instructions are found** is meaningless for the UDC, since the rules for this combination are general and are not stated under any specific number.

Furthermore, the base number is only *78 Music*, the remainder are numbers sourced from different tables, so for each number it would be necessary to state the table identification. Combinations consist of *78 Music* (main number from main table), *.071 Musicians* (special auxiliary table), *-056.45 Talented persons* (common auxiliary table - persons) and *(=411.16) Jews* (common auxiliary table - ethnic grouping) which is actually built from the parallel division of the common auxiliary table of languages *=411.16*. The suggested approach to coding is inadequate for UDC number coding, quite apart from being limited in supporting the necessary functions.

Also, 'span' numbers are coded in the main heading field. It was suggested that other relationships such as colon (:), plus (+) and subgrouping (::, []) should be coded as control field **$4 Classification syntax** to be used within other blocks of data : specifically **2xx** - heading block, **3xx** - history notes, **7xx** - index term fields. This means that whenever two numbers are combined with something other than / (span) they would be connected with tags from the **$4** schema.

If one sets aside the advantages of having any classification format compared to having none, one should notice that the UNIMARC classification format has missed the opportunity to become the long-awaited solution for many libraries using UDC and other general or special synthetic classification schemes. Coming after MARC 21, this format had the opportunity to provide an extended and improved way of managing different classification systems. Since its publication on the IFLA

Website, the format is still in its draft form and is awaiting proper revision.

6.3.5.3 Proposal for improvement of UNIMARC classification format

After the first draft of the format was analysed and considered unsuitable, a suggestion for a different, fully structured number heading field was proposed to the Working Group on Classification Format in 1999 and again in 2000 and 2001.[84] The main characteristic of the suggestion was that the approach to handling synthesised numbers should be changed by allowing the main classification-heading field to be fully coded. So, while MARC 21 offers this approach for notation from the main tables (main numbers) synthesised by a 'span', UNIMARC would provide this functionality for all types of notation combinations. This means that all synthetic notations would be given the same functionality and they would have subfields allowing for their direct manipulation, access and global change.

In addition, the fact that this format is meant to be used for both enumerative and synthetic classifications and because it may be one's deliberate choice to treat synthesised numbers as a single number, it was necessary to make the format easy to use for the management of enumerative classifications such as DDC and LCC or those requiring a simple approach to management data. The only way to address this is to offer two parallel fields for classification heading:

250 Class Number for enumerative classification systems
251 Class number for analytico-synthetic classification systems

The main concern of the proposal for improvement is to provide machine readability of each semantically distinct part of a classification number on the level of the classification heading, but based on specific criteria that can be determined across different synthetic classification schemes.

The systems analysed for this purpose were UDC and BC2 whose structures correspond or are similar to many faceted and synthetic special classification

[84] Maria Inês Cordeiro, Gerhard Riesthuis and Aida Slavic devised the proposed changes discussed in this section in 2000.

schemes such as, for instance, Forest Decimal Classification (FDC IUFRO). It was decided that separate subfields should be provided for the following two types of number components:

- elements having exactly the same value (string of numbers) and having the same meaning irrespective of their combinations, which can be combined with any other element without restriction (e.g. common auxiliaries in the UDC)
- elements consisting of exactly the same value (string of numbers) bearing the same meaning within a class or set of classes, and in case there is a need to distinguish such parts syntactically for the purpose of sorting (e.g. special auxiliaries in the UDC, such as 01, -1/-9, and ' in compound numbers)

The elements of combinations, for which there will be no separate subfields, would be indicated and described in field **675 Synthesised number components**:

- elements that bear the same meaning across either all classes, or a set of classes, but are not represented with the same value (string of numbers). For instance, these would be typical for the notation of fundamental vocabulary facets such as techniques, methods, operations, agents, materials etc. that although they may exist across different subjects have a different notation
- elements bearing the same meaning across either all classes or a set of classes, with the exact value (string of numbers) that are prefixed or changed in a compound number and cannot be reached by post coordinating the parts. This is the case with parallel division / languages

With respect to relationship symbols, it was decided that they should all be coded in **control field $4** which will allow for their flexible use across any kind of number combination. Apart from those used in the UDC, the set of relational symbols could also be extended to include phase relationships as they are defined in BC2. The control field would also provide tagging for UDC special auxiliaries and alphabetical extension as shown in the Table 6.3.

Table 6.3: Proposal for coding relationships in UNIMARC Classification Format schema

$4 Classification syntax and attribute code		
Type of Relationship	**Value**	**Symbol (UDC)**
Coordination	a	+
Simple relation	b	:
Order-fixing	c	::
Subgrouping	d	[ ]
Span in auxiliary numbers	e	/
Phase relations (BC2): Relations with other subjects. Comparison	f *followed by the 2-letter code adequate:* f6T or f9C	
Relations with other subjects. Exposition	f6U or f9E	
Relations with other subjects. Author's viewpoint	f6V or f9F, f9G	
Relations with other subjects. General influence. Effect	f6W or f9H	
Relations with other subjects.Influence by another subject	f6X or f9J	
Relations with other subjects. Influencing another subject	f6Y or f9K	
Type of special auxiliary (UDC)		
Special auxiliary table using .0	g	.0
Special auxiliary table using -1/-9	h	-1/-9
Special auxiliary table using ' (apostrophe)	i	'
Use of Auxiliary Table Numbers as Main Numbers	j	

The most important difference in this proposal is, however, the syntax of the alternative classification heading field **251 Class number for analytico-synthetic classification systems**. The subfield schema suggested is comprehensive enough to support both BC2 and UDC. All subfields are optional and repeatable.

Some compromise was made so that the tagging scheme, shown in Table 6.4, agrees as far as possible with the tagging scheme in existing field **250** which is reserved for classification numbers that are used for unstructured classification headings, which is why subfields **$a** and **$c** are exactly the same.

Table 6.4: Proposal for encoding schema for class number in UNIMARC

TAG	FIELD DESCRIPTION
251	**Class number for analytico-synthetic classification systems**
$a	**Main classification number or beginning of span**
$c	**Ending classification number of span**.
$d	**Special auxiliary numbers** Special auxiliary numbers (e.g., in the case of UDC, taken from the hyphen series -1 / -9, the point-nought series .01 / .09, or apostrophe series '0 / '9. In the case of BC2 this subfield could be used for the remaining special auxiliary tables in Language and Literature).
$e	**Common auxiliaries for Time** Auxiliary number representing a time period (e.g. taken from UDC Table Ig – Common Auxiliaries of Time, or BC2 The Auxiliary Schedule 4 – Time)
$f	**Common auxiliaries of race, ethnic grouping and nationality** Auxiliary classification symbol representing ethnic group (e.g. taken from UDC Table If – Common Auxiliaries of Ethnic Grouping and Nationality, or BC2 Auxiliary Schedule 3A – Ethnic Groups).
$g	**Common auxiliaries of Place** Auxiliary classification symbol representing place (e.g. taken from UDC Table Ie– Common Auxiliaries of Place or BC2 The Auxiliary Schedule 2 - Place). In the case of a span the second $g should be preceded by $4e.
$h	**Common auxiliaries of Form** Auxiliary classification symbol representing form (e.g. taken from UDC Table Id – Common Auxiliaries of Form, or BC2 Auxiliary Schedule 1 - Common Form Subdivisions).
$i	**Common auxiliaries of persons and personal characteristics** Auxiliary classification symbol for persons (taken from UDC Table Ik -05 – Common Auxiliaries of General Characteristics: Persons, or BC2 Auxiliary Schedule 1A - Persons).
$j	**Common auxiliaries of Languages** Auxiliary classification symbol representing language (e.g. taken from Table Ic – Common Auxiliaries of Language, in the case of UDC or taken from BC2 Auxiliary Schedule 3 - Languages).
$k	**Common auxiliaries of properties** Auxiliary number for properties (e.g. taken from UDC Table Ik-02 – Common Auxiliaries of General Characteristics: Properties).
$l	**Common auxiliaries of materials** Auxiliary number for materials (e.g. taken from UDC Table Ik-03 – Common Auxiliaries of General Characteristics: Materials).
$m	**Common auxiliaries relations, processes and operations** Auxiliary number for materials (e.g. taken from UDC Table Ik-04 – Common Auxiliaries of General Characteristics: Processes).
$n	**Symbols from external sources** Notation taken from a source external to the classification system (e.g., in the case of UDC, those denoted by * (asterisk) as prefix, according to Table Ih – Subject Specification by Notations from Non-UDC Sources).

The following examples show differences in coding synthesised UDC numbers. The first example shows the coding of a combination of main number and common auxiliary number using subfields of field **251** (coding according to scheme in Table 6.4).

004.421(075.8)=111Texts for university, higher education - in English - about algorithms for program construction	
251 ##$a**004.421**$h**075.8**$j**111**	INSTEAD OF: 250##$a**004.421(075.8)=111**

The following example containing the colon combination of two numbers shows how control field **$4**, from Table 6.3, is used to specify a colon using coded value **b.**

32:34"18"(73)Relationship between politics and law the United States of America in the 19th century.	
251 ##$a**32**$4b$a**34**$e**18**$g**73**	INSTEAD OF: 250 ##$a**32:34"18"(73)**

The last example shows a complex number with a range of numbers appearing in both the main number and common auxiliary. The range in the main number is coded using subfield **$c** which was already introduced in MARC 21 and UNIMARC and is not changed for the purpose of compatibility. The range in the auxiliary number of time *"15/16"* is, however, coded using control field **$4** and the code for 'span' **e**

011/016"15/16"(091) History of bibliographies in the 16th and 17th century.	
251 ##$a**011**$c**016**$e**15**$4e$e**16**$h**091**	INSTEAD OF: 250 ##$a**011**$c**016"15/16"(091)**

This approach in coding, if adopted, would allow the access, retrieval and management of all component parts of synthesised UDC numbers. Control over the facets of vocabulary such as place, time, form, and language is complete and independent of special symbols used in the classification notation and allows proper sorting as well as global changes of each element in all combinations into which they are built. This allows full control and tracing of semantic inheritance and an advanced way of search expansion. Although this approach may seem more complex, the time spent is exactly the same as that for coding using field **665 Synthesised number** component in the current field – but with greater advantages and control. If this approach in coding were accepted for UNIMARC classification authority files, then the same change could be easily extended to field **675** of the UNIMARC Bibliographic format, which is reserved for UDC.

6.4 Summary of UDC authority control

The requirements for UDC data retrieval and the classification interface (Chapters Four and Five) stressed the importance of processing UDC data in a way that enables access and control over structural elements of pre-synthesised UDC numbers, semantic relationships and linking numbers to natural language. In this chapter, the specific requirements for processing and control over these elements of UDC data were summarised.

This research confirms that the coding of individual elements of pre-synthesised class marks and the coding of hierarchical relationships, independent of the UDC original symbols and notation, is paramount. In addition, the link between UDC notation and natural language terms needs to be provided for, maintained and controlled for both caption and caption-independent search terms. For this to be achieved, UDC data have to be managed as an authority file, external to the actual bibliographic database to which they are linked. Apart from authority control, access control, validation, automation of semantic linking and global changes of class marks, the classification authority data can be used for classification sharing and exchange.

The re-use and exchange of classification authority files across systems, as well as the ability to search distributed databases using UDC, will depend on the standardization of an authority file format which will ensure its use in a greater number of IR systems. In the bibliographic domain this can be easily achieved if the UDC authority file is made in one of the MARC formats for classification authority data, and two of these formats have already been developed. This is a large step forward in comparison to the previous poor treatment of classification in library systems.

MARC 21 and UNIMARC classification formats, which are analysed in this chapter, however, show that neither provides data coding that would facilitate the required UDC data processing. This means that UDC, if used in such a format, would not be used to its full extent. A suggestion for improvement of UNIMARC Format for Classification Data, described at the end of this chapter, resolves the majority of problems present in the existing format draft. The further development and adjustment of UNIMARC is important, not only for libraries using UDC, but also for those who own and maintain the UDC. Classification data that is distributed has to be in a format compliant with the standard format used in the bibliographic domain. In other words, a mapping between UDC MRF file data structure and UNIMARC or MARC 21 format has to be possible.

Up to this point in the research, it has been established that the use of UDC in IR depends on appropriate data processing and that work has already been undertaken in the bibliographic domain to embrace more advanced management of classification data. It is also apparent that MARC formats for classification data, even if they did fulfil all these requirements, would still need to be compatible with UDC source data in terms of structure and support for data processing.

In other words, the conversion of UDC MRF data to a MARC authority data format should be not only possible but straightforward and the UDC data should normally be distributed in this format to licence holders in libraries. In the next two chapters the focus will be on the suitability of the existing UDC data as it is structured and stored in the UDC MRF, for the task of the management, maintenance, revision and distribution of the UDC.

CHAPTER SEVEN: UDC USE, MAINTENANCE AND DISTRIBUTION

7.1 Introduction

In 1993, UDC was the first general classification system to be distributed to users as a database, with associated database exports, and this was in conjunction with other notable changes in classification management at the time. Ever since, the database file has been the only form in which the system has been maintained, updated and distributed. The original plan of the UDC Consortium[85] was that with a new revision policy, UDC should undergo complete restructuring within approximately ten years (Task Force for UDC system development, 1990: 7). In the period 1993-2005 there were ten annual updates and although UDC was not completely revised, many changes took place and these were distributed to publishers and users as they were introduced.

During this ten-year period, the change of ownership, management, revision process, product, distribution, buying conditions and price must have had an impact on the UDC user community and the way in which the classification was used.[86] In order to shed some light on the reality of UDC use worldwide this chapter starts with a UDC user survey conducted in 2004. In addition to this, a bibliographical survey of UDC translations which existed in 2004 was conducted to clarify not only the total number and type of translations but to give an account of how the number of new translations, bilingual editions and electronic editions relates to the automation of the UDC and change of its ownership.

---

[85] Henceforth the Consortium of publishers that owns and manages the UDC will be referred to as "the Consortium".

[86] Information on management of the UDC has been obtained during research and practical work undertaken for the UDC Revision Committee. This comprises both schedule development and maintenance, and update of the UDC file for the period 2000 to 2001. The *Extensions & Corrections to the UDC* (E&C) is the UDC Consortium official publication, used to communicate the most important decisions and changes to the UDC users and all the issues in the period 1993-2004 were consulted. Additional sources included FID and UDC Consortium internal documentation such as *Task Force for UDC System Development: final report (1990)*, *Guidelines for the creation of a standard version of UDC (1991)* and *Master Reference File Manual (2003).*

The central part of this chapter looks into the relationship between schedule automation and UDC management, and the revision and distribution of its products to end-users. The UDC database format was created based on the presentation of the UDC within the printed edition and, as pointed out by Strachan & Oomes (1993, 1995), in 1993 there was still a large task ahead improving the data structure, quality and consistency. The UDC MRF is sold in the same form both to the classification publishers and users (i.e. those who use UDC in indexing and information retrieval). Central to this chapter is how and to what extent the UDC maintenance and revision processes have continued this initial work in improving UDC content and its presentation and to what extent the database is used and how, in supporting the revision process and user needs.

7.1.1 Worldwide UDC user: survey in 2004

The UDC user survey, as explained in the Research methodology (Chapter Two, section 2.2), combines the following techniques: user interview via e-mail, bibliographical literature research, and UDC user data obtained from other sources.[87] The objective of the simple interview technique applied was to collect and verify basic data on UDC use worldwide, via e-mail contact with libraries and information services, in as many independent countries and overseas departments or dominions as possible.

Based on preliminary data (Appendix 8.4, Table 8.1) and research of the literature, the choice of e-mail contacts was set, whenever possible, to include cataloguing departments of national, academic and special libraries. For many regions, especially in Africa, Asia, South Pacific, Central and South America, in which UDC was more likely to be used in documentation and information centres, these were included in the survey as well as library associations, LIS schools and archives. Within these parameters, the choice of institution to contact was random, and the less information on the country's classification practice was available, the more e-mail contacts were attempted. The interview data were then compared and

---

[87] The UDC Consortium provided a list of countries holding licence agreements, and the Spanish Consortium member AENOR provided a list of their customers (Appendix 8.4. Table 8.1). The Belgian (CEFAL) and Russian (VINITI) Consortium members did not respond to the request for data.

compiled with research of the literature. It was beyond the scope of this survey to attempt to establish the number of institutions or exact percentage of libraries using UDC in any given country.

Earlier user studies. In the period 1990-2004, the Consortium did not publish or disseminate any data on UDC users worldwide and no official estimate of the total number of countries or institutions using UDC was released to the public in this period.[88] The last UDC user survey conducted is mentioned by the *Task Force for UDC System Development*'s report published in February 1990, which refers to the user survey conducted by FID in 1989 to which they had access.[89] The following information provided by D. Strachan was, however, the only information on this user study found in 2004: *"The user survey took the form of a printed A4 questionnaire, and was an initiative of the recently formed UDC Management Board. I cannot recall the exact timing, but I think that preparation was begun sometime in 1988 and collation of the returned data was completed before the end of 1989. The purpose was both to get information on who used the UDC and in what way, and to invite their ideas on deficiencies and priorities. The questions were in parallel texts in English, German and French. Distribution was both direct (using mailing lists e.g. of subscribers to E&C) and indirect (with the help of FID national members, national UDC committees, and publishers of the various editions). For example we took advantage of the distribution of the English Medium Edition in 1989 by persuading BSI to enclose a questionnaire with each completed order.*

The replies in the returned forms were collated as a Dbase database - each question having its own field. I cannot remember how many forms we sent out, nor how many were returned completed. I prepared a summary report for the Management Board who decided that the returns were not sufficient to merit further

---

[88] The last **published** survey on UDC use was on UDC users in the UK in 1979/1980 by Hindson (1981) who established that there were 640 libraries using UDC in UK and Northern Ireland (out of 2, 895) at the time.

[89] The reference in the Task force report in relation to the survey consists of the following two sentences: "*... it is strongly recommended that the 108 replies to FID's questionnaire which expressed willingness to co-operate in revision work should be actively followed up...*" (page 8) and "*Full access to relevant papers has been provided by FID to the chairman and all members of the Task Force, including the summary of results from the questionnaire issued by the FID Classification Department in 1990*" (Task Force for UDC system development, 1990: 11-12)".

action beyond making them available to its Task Force for UDC System Development which was doing its work at that time. No report was published." (Strachan, 2004)

When enquiries were made in 2004, the Consortium was unable to provide any information on the results, not having the access to the questionnaire material, which was deposited in the FID archive and was inaccessible due to major building works at the Royal Library in The Hague, at the time of this research.

The appendix of the *Task Force's* final report in 1990 contains the *Questionnaire* and *Summaries of results from individual investigations* with three reports on research that was undertaken in 1989. The first report is the investigation of twelve institutions using UDC, eight of which were in the UK. Four overseas institutions included were stated as: a Japanese library, an Eastern European National Bibliographic Agency, a Nigerian University and an African Documentation Centre (the names of countries and location of these institutions are not given). The second report is an account of the interviews with chief librarians in nine libraries from Denmark, Finland, Norway and Switzerland (the name of the institutions is not provided). The third report is a summary answer in the form of a questionnaire from six Austrian libraries using UDC (the full name of the institutions are given). The criterion for the selection of 27 institutions, as representatives of UDC users, was not explained in the reports. All three investigations seemed to be based on a questionnaire (13 questions) that focused on the way in which each institution used UDC and their expectations from the scheme. Reports indicated that some of the interviewees were considering a change to another classification scheme in the future (Finland) and some to abandon using classification all together if this would pose a problem for automation, as it was the case in Austria (Task Force for UDC system development, 1990).

The only published data on UDC use worldwide in the 1990s was an estimate in an Estonian article (Nilbe, 1997). The author mentioned 60 countries and 100,000 institutions using UDC based on literature from the 1980s.[90]

---

[90] The FID survey in 1968 indicated 100,000 users mostly in the USSR, European countries, Latin America and Japan (Gilchrist, 1992). The broad library network in the former USSR consisted of 300,000 libraries and the UDC was used as an obligatory system in scientific-technical libraries only (Sukiasyan, 1988:69).

7.1.1.1 Procedure

One feasible goal for the 2004 user survey was set to establish the number of countries using UDC and their assignment to one of the following simple categories:

Countries A UDC is used in the majority of libraries within the country and is the dominant classification system

Countries B UDC is used in some type of libraries and in these countries it is used alongside other dominant classification system(s). UDC in these countries is a well known classification system and is most likely to be taught in LIS schools.

Countries C In these countries UDC is rarely used compared to other systems and rarely taught in LIS schools.

The difference in the number of institutions using UDC implied by these three bands is significant and is relative to the size of a country. Countries in band A are likely to have several thousand libraries using UDC as these may include public, academic, special and school libraries. Use of the UDC in these countries is quite stable and libraries do not change to other systems very often.

Countries in band B, depending on their size, are likely to have anything between twenty and several hundred libraries using the UDC. The number of institutions may be proportionally greater in large countries such as the Russian Federation, but no estimate can be made without further research. This band is where most changes occur over time, with respect to the number of institutions using UDC. In this band of use, libraries may migrate from UDC to more dominant systems, as is occasionally the case in Western Europe. Also, libraries may change to UDC as is the case in countries from the former USSR. Countries in band C, even large countries, are not likely to contribute much to the total number of institutions using UDC but they are a significant indicator of the viability of UDC as an international scheme.

In the process of the research, however, it was established that for many countries it would be more accurate to use categorisation in between the established bands A/B or B/C, to denote A with a tendency towards B or B with a tendency towards C. Throughout the e-mail interview, an attempt was made to learn about any published national library survey statistics but this was rarely available. The general framework of the interview is shown in the Appendix 8.1 and samples of answers illustrating various level of information quality are provided in Appendix 8.3.

There were on average 6 mails sent to each of the 208 countries (not counting 187 mails that were returned as undelivered). Contact was established with 153 countries (73%), and an average of 3 responses was obtained per country. E-mail correspondence was then established to obtain information or further contact details and clarify/verify existing information. Summary statistics of responses are given in Table 7.1 and a list of countries surveyed is provided in Appendix 8.4, Tables 5, 6, and 7.

Table 7.1: Survey 2004 coverage

	TOTAL NUMBER OF COUNTRIES SURVEYED	STATUS UNKNOWN	RESPONDED	
			NOT USING UDC	USING UDC
EUROPE	45	11% (5)	0	89% (40)
ASIA	50	30% (15)	22%(11)	48% (24)
AFRICA	53	36% (19)	19%(10)	45% (24)
NORTH AND CENTRAL AMERICA	30	20% (6)	50% (15)	30% (9)
SOUTH AMERICA	13	15% (2)	0	85% (11)
AUSTRALIA AND SOUTH PACIFIC	17	47% (8)	29% (5)	23% (4)
			73% (153)	
	208	26% (55)	20% (41)	54% (112)

In parallel with the e-mail interview, data on UDC use in national bibliographies were checked in *An annotated guide to current national bibliographies* and related work by Bell (1998, 2001). Also, countries with less known library practice were checked in the *Encyclopedia of library and information science* and in library journals related to cataloguing and classification. In addition, the following data were collected (Appendix 8.4, Table 3):

- The Consortium data: the number of countries from the Consortium licence agreements; the number of countries from UDC users' web pages, and AENOR non-European UDC products buyers
- IUFRO Global Forest Decimal Classification, 2003 survey of forest libraries in 27 countries (19 using UDC)
- Countries of participants on the udc-forum mailing list

7.1.1.2 Findings on UDC use

Complete survey data with details of each of the 112 countries using UDC is given in Appendix 8.4. Table 7, and this includes the following information: the type of use, languages in which the UDC is used and whether it is used in the national bibliography. According to Bell (1998, 2001), the UDC is used for subject organization of national bibliographies in only 28% (32) of countries and these are mostly in Europe (24).

This research was focused on coverage and collecting elementary data of UDC use, which is why the sources for each country were only checked and re-checked until the answer was reasonably clear. The total amount of information verified for each country was graded for confidence or reliability as 1 - very confident, 2 - fairly confident or 3 - double-checking desirable. In some cases the information was confirmed by several correspondents, in others only one person provided the data. The quality of response also varied as shown in Appendix 8.2. The lowest reliability (3) is more often found in correspondence with African, Asian and South American countries with which it proved harder to find and establish contact.

Results summarised in Table 7.2 show that there are 112 countries where UDC is used and these are distributed across bands showing the largest percentage of countries: bands B-BC (40%).

Table 7.2: Count of countries using UDC according to the band of usage

COUNTRIES USING UDC	A or A/B			B or B/C			C
EUROPE (40)	22 (55%)	A A/B	19 3	12 (30%)	B B/C	8 4	6 (15%)
ASIA (24)	4 (17%)	A A/B	1 3	8 (33%)	B B/C	5 3	12 (50%)
AFRICA (24)	8 (33%)	A A/B	3 5	11 (46%)	B B/C	4 7	5 (21%)
NORTH AND CENTRAL AMERICA (9)	—			3 (33%)	B B/C	2 1	6 (67%)
SOUTH AMERICA (11)	—			8 (73%)	B B/C	3 5	3 (27%)
AUSTRALIA AND SOUTH PACIFIC (4)	—			3 (75%)	B B/C	1 2	1 (25%)
TOTAL (112)	34 (23 A + 11 A/B) 30%			45 (21 B + 24 B/C) 40%			33 30%

Within the total number of countries using the UDC it appears to be the dominant classification system in 30% countries (A-AB), namely Europe (23), Asia

(3) and Africa (8). A third of countries (30%) are in band C in which the number of institutions using UDC is very low.

With regard to languages, 27% of countries have the UDC translated and used in their national language (Table 7.3). Logically, these are countries where UDC is the dominant system.

Table 7.3: UDC use in world languages

LANGUAGE	COUNT OF COUNTRIES	USAGE BAND	COUNTRIES[91]
ENGLISH	**31** (26 English + 1 E/French 1 E/Finnish 1 E/Indonesian 1 E/Hebrew 1 E/Nor)	**B**	B: Australia, Finland, Norway, Ghana, South Africa, India, UK B/C: Ireland, Canada, Uganda, Indonesia, Israel, New Zealand
		C	Kenya, Nigeria, Tanzania, Bangladesh, Bhutan, Singapore, Sri Lanka, Fiji, Greece, Iceland, Malta, Trinidad & Tobago, USA, Sudan*, Malaysia*, Syria*, Barbados*, Jamaica*
FRENCH	**17** (16 French + 1 F/Dutch)	**A**	A: Congo DR, Togo A/B: Belgium, Algeria, Burkina Faso, Mali, Senegal
		B	B: France, Luxembourg, Tunisia B/C: Benin, Libya, Madagascar, Morocco, Niger, French Polynesia
		C	Cameroon
GERMAN	**3** (2 + G/French/English)	**A**	A/B: Switzerland
		B	B/C: Germany
		C	Austria
PORTUGUESE	**7** (6 P + 1 P/Chinese)	**A**	A: Portugal, Angola A/B: Mozambique, Macao (China)
		B	B: Guinea Bissau, Brazil B/C: Cape Verde
RUSSIAN	**9** (7 + 1 R/Kyrgyz 1 R/Azeri	**A**	A/B: Armenia, Azerbaijan,
		B	B: Kazakhstan, Kyrgyzstan, Uzbekistan, Belarus, Russian Federation
		C	Tajikistan, Turkmenistan*
SPANISH	**15**	**A**	Andorra, Spain
		B	B: Cuba, Argentina, Uruguay B/C: Costa Rica, Bolivia, Chile, Colombia, Ecuador
		C	Guatemala, Mexico, Paraguay, Peru, Venezuela
OTHER LANGUAGES	**30**	**A**	A: Albania, Bosnia & Herzegovina, Croatia, Czech R., Estonia, Georgia, Hungary, Latvia, Lithuania, Macedonia, Moldavia, Poland, Romania, Serbia & Montenegro, Slovakia, Slovenia, Ukraine A/B: Bulgaria
		B	B: Denmark, Sweden, Japan, B/C: Vietnam, The Netherlands, Surinam
		C	China, South Korea, Taiwan, Turkey, Faeroe Islands, Italy

[91] Asterisk (*) denotes countries with low UDC usage i.e. with only a very few institutions using UDC.

English is the most frequent language and is used in 31 countries (28%). However, all of these countries are B or C band of usage and the number of institutions using UDC in these countries is not very significant. French is the next most used language (17 countries) followed by Spanish, used in 15 countries, and both of these languages are used in some countries where UDC is the main classification system. Russian and Portuguese are also used across countries with a large percentage of UDC users.

Europe. The research in 2004 confirms that Europe has the most significant concentration of UDC users (40 countries), with 22 countries (55%) in the A and AB band of use (mostly East European countries). The dissolution of the former Yugoslavia and Czechoslovakia augmented the total of countries but the number of actual users of the UDC was not affected by this particular change in the 1990s. The dissolution of the USSR, however, increased both the number of countries and the number of users. The reason is that countries such as Estonia, Latvia, Lithuania, Ukraine and Moldova started to use the UDC as the main classification system, replacing the Soviet Library and Bibliographic Classification (BBK) (Nilbe 1997, Gobytė, 2000). Research into band B shows a number of countries with a long history of the UDC use.[92] In some European countries, however, the number of UDC users was in decline in the 1990s and this was confirmed for Austria, France, Germany[93], Italy, the Netherlands and the UK. The reason for this is the pressure to reduce cataloguing/classification and national networking costs by migrating to the country's main classification system. In Europe the UDC is used in national languages with the exception of English schedules that are also used in Scandinavian countries, Greece, Malta[94] and Switzerland, and French translations in Belgium, Luxembourg, Switzerland (Table 7.3)

Asia. The survey did not manage to collect data from 15 out of the 50 Asian

---

[92] For Norway (Anjer, 2004) the number of users is estimated to be up to one hundred libraries and Benito (2001) confirms a similar number for Sweden. One third of around 150 special libraries in Denmark were using UDC in the 1980s and the situation does not seem to have changed significantly (Pedersen, 2004).

[93] A summary of objections that Austrians and Germans have in relation to the UDC was sent to the UDCC in 2000 and it specifies the complaints about a lack of support from UDCC, inconsistency and incomparability of old and revised data, unsuitability of the UDC data format and tools, and the high price of the MRF licence (Wagner, 2000).

[94] English is the 2nd official language in Malta, after Maltese.

countries surveyed (30%) as seen in Table 7.2. Collected data shows that 24 are UDC users (48%) and that among these countries 50% belong to band C. Mainly due to the dissolution of the USSR and wider adoption of the UDC in countries such as Armenia, Azerbaijan, Georgia, Kazakhstan etc., 4 Asian countries (12%) are A-AB and 8 are the B-BC type of user (33%). The only statistics on UDC use available for an Asian country were from research conducted in 1989 that shows that the UDC is the third most used classification in India, and is used in 15% of surveyed libraries (Satija, 1993).

The decline of UDC use in Asia was hard to investigate as the general usage level is low and literature is not abundant. Qiyu, & Xiangsheng & Dongbo (1996) mention a wider use of the UDC in technical libraries in China in the 1960s than is the case in 2004. It is significant that the use of the UDC in some Asian countries that have been known to use UDC in the past was not confirmed in 2004. Such is the case with the former FID members Kuwait and Lebanon, and also with Iran and the Philippines.

Africa. The 2004 survey confirmed Diongue-Diop's (1992) findings that the use of UDC in Africa is present across francophone Africa. The survey also confirmed the use of UDC in English speaking Ghana, Uganda, Kenya, Nigeria, Tanzania, South Africa and Sudan and Portuguese speaking Angola, Cape Verde, Guinea Bissau and Mozambique. In total 45% of African countries are using UDC. Data are still missing for 19 African countries (30%) and much of the data uncovered are in need of further verification. Eight countries (33%) have UDC as the dominant system and 11 are in band B-BC (45%). South Africa is one of the representatives in this band and according to the directory of Southern African libraries from 1983 there were 95 mostly special libraries using UDC (Van der Walt, 2004). Out of the total number of African countries, five have only a few libraries using UDC (22%). The decline in use was reported for Nigeria and Tunisia, and there appear to be no libraries left using UDC in Egypt.

North and Central America. This part of the world has never had many UDC users and this was confirmed with only 9 out of 30 countries (30%) in the area using UDC and 15 (50%) confirming that there is no library using the system. Also, 6 of the countries using UDC in this region (67%) are in band C with only a small number of institutions (mostly special libraries). The remaining 33% are in band B-BC. Dominant languages in which the UDC is used are English in Barbados, Canada, Jamaica, Trinidad & Tobago and the USA, and Spanish in Cuba, Costa Rica, Guatemala and Mexico. No information was found on the change of UDC use in the 1990s.

South America. UDC is used mainly in special and, occasionally, in academic libraries and documentation and information centres (often banks) in 11 South American countries (85%). A larger proportion, 73% of countries (8), is in band B and the remaining 27% in band C. Although dominant languages are Spanish and Portuguese the interest for English/Spanish/Portuguese electronic editions was noted and this is related to the information library network that connects Central and South America. Information on the UDC user community in South America is scarce and more research would be needed to acquire any knowledge about the change in the number of UDC users in the 1990s.

Australia and South Pacific. Only 4 countries (23%) in this region use the UDC. For 8 countries (48%) the status is unknown and 5 were able to confirm that there was no library using UDC (29%). Two main UDC users in this area are Australia and New Zealand and they are classed in usage band B-BC while a smaller number of users are scattered across Fiji and French Polynesia. In Australia, the UDC is used primarily in special libraries. Harvey (1998: 215-216) confirmed that there is no published survey as such but that there was still enough isolated information to conclude that the use of the UDC in Australia was significant. He mentioned that in Victoria in 1980s there were 19 institutions using UDC in addition to several CSIRO (Commonwealth Scientific and Industrial Research Organisation). Harvey also mentioned a 1988 survey of libraries in Western Australia that reported 30% of special libraries were using UDC. Migration from UDC to DDC is supposedly planned in French Polynesia.

7.1.2 UDC translation into world languages: 2004 bibliographic survey

The goal of the UDC bibliographical survey in 2004 was to provide data on the total number of languages in which any kind of UDC schedules have appeared.[95] The idea was that an overview of the date and type of these editions in the period before and after 1993 could be taken as one of the indicators of the global strength of

---

[95] The survey looked for the last reported translation in each language.

the UDC and the vitality and beneficial influence of the changes in UDC management and data accessibility. The main obstacle in collecting data on UDC editions is the flexibility of the UDC system, which allows users to adapt the system, extend or change any part of the schedule, or combine it with some other classification. It was therefore decided to collect information only on full, medium, abridged and special editions last reported for any given language. The textbooks to the UDC were not analysed.

The preliminary study showed that there are four sources of information published over thirty years that contain useful information. Chronologically, the first source is the textbook *La Classification Décimale Universelle* (Dubuc, 1973). The second is the summary table of published editions in *Automation and the UDC*: 1948 - 1980 (Rigby, 1981). The third resource is a bibliography proper on UDC editions covering the status in 1981, published in 1982 by FID (A Bibliographical Survey of UDC Editions, 1982). After 1982 there was no cumulative bibliography and no new research on translation of the UDC was reported. The only source was occasional information on translations as they appeared in the bibliography and text of *Extensions and Corrections to the UDC* during the period 1993-2004.

Procedure. When compared, the four above-mentioned sources provide information on different editions appearing in a total of 28 languages (Appendix 9.1). This information was taken as a starting point and countries in which these translations appeared were first targeted when looking for more recent translations. A selection of other countries was then made from various reports on national bibliographic control or on the use of the classification in bibliographic services worldwide (Bell, 1998, 2001; Diongue-Diop, 1992; Haarala, 1991; MacMillen, 1992; Satija, 1993; Jarvis, 1993; Nilbe, 1997). A final selection of countries contained all independent countries from the former Soviet Union, some in South East Europe (e.g. Albania, Bulgaria, Greece, Malta, Turkey) and some in Asia (e.g. China, Indonesia, Korea, Malaysia, Mongolia, Vietnam).

References on UDC editions were collected from Web OPACs, where available, and wherever language and script were not a problem. Search terms in various languages were collected either from other national Web pages, from library schools or from the literature - so when a catalogue was in a script other than roman, search expressions were copied and pasted into the OPAC search box. Alternatively, a search was performed using class number *025.4* or *025.45 Decimal classifications* meaning the same in both Dewey and UDC catalogues. The resulting titles would simply be copied and used to perform searches in other catalogues in the same country. In parallel, national, academic, special and public libraries were contacted by e-mail and asked about their use of the UDC and translation into their languages. Finally, the Consortium was approached for additional information on the number of languages for which a translation licence was requested in the period 1990-2004 and this information when checked confirmed existing data on the number of languages.

7.1.2.1 Findings on UDC translations

The result of this research, as shown in the Table 7.4 (next page), confirmed that the UDC was translated into 39 languages and bibliographic references were collected for 37 languages (see Appendix 9.3). Comparisons between previous reports and the 2004 study show a significant growth in the numbers of languages (Appendix 9.2, Table 9.5). There are two reasons for the large discrepancy in the number of languages (16), compared with the last bibliographic survey published by the FID in 1982. The first reason is that some new translations are encouraged by political changes in Europe. The second reason is that the FID bibliography was limited, to start with, by the number of translations recorded by the FID office at the time.

In effect, only 7 translations in 'new languages' appeared during the period 1993-2004 and this happened mainly because of the political change in Czechoslovakia, the USSR and Yugoslavia (Croatian, Estonian, Georgian, Latvian, Lithuanian, Slovakian, Ukrainian). A Chinese translation in 1997 was published in Macao, after the country was reunited with China. Up to this point Macao was using Portuguese editions. Many different Chinese translations, however, existed before this. Attention should also be drawn to translations in the 1980s and 1990s in Albanian, Bulgarian, Catalan and Vietnamese most of which were not related to any particular political change at the time.

Of the total number of translations reported here, there are 12 whose last reported date of publication was before 1990. Four were published in 1980-1990 and the remaining 8 editions are all dated before the 1980s and could be said to have mainly historical value. These old editions (Azerbaijani, Hebrew, Icelandic, Indonesian, Kyrgyz, Norwegian or Turkish) are included as they represent countries where UDC is still used.

Table 7.4: Last Reported Editions According to the 2004 Survey[96]

LANGUAGE	LAST DATE	FULL	MEDIUM / MRF	ABRIDGED	SPECIAL	ELECTRONIC	VERIFY[97]
Albanian	1997s	—	—	1993-1997s	—	—	C
Azerbaijani	R[98]	—	—	1960?	—	—	C
Bulgarian	1985	—	1985	—	—	—	O
Catalan	1982	—	1982	—	—	—	C; O
Chinese	1997	—	—	1997	—	—	C
Croatian	2004-	—	2004*	2003	—	—	C; E&C
Czech	2002	—	1994-1995	—	—	2003W; 2003C	C; E&C
Danish	1995	—		1986-1995	—	—	C; O
Dutch	2003	—	2003	—	—	—	E&C
English	2003	—	1993	2003	—	2002W; 2003F[99]	E&C
Estonian	1999	—	1999	—	—	—	E&C
Finnish	1991s	—	—	1983-1991s	—	—	O
French	2004	—	2004	2001	—	—	E&C
Georgian	1998	—	—	1998	—	—	C
German	2002*	—	1978-1985	2002*	—	—	E&C
Hebrew	1969	—	—	1969	—	—	C
Hungarian	1992	—	1992	—	—	—	E&C
Icelandic	1953	—	—	—	1953	—	C
Indonesian	R[100]	—	—	—	1970s		C
Italian	2000	1972-1985	1974	1987	—	2000 (abr)	O
Japanese	2004	—	1994	—	—	2004C Jap./En.	E&C
Korean	1973	—	—	1973	—	—	C
Kyrgyz	1961	—	—	1961	—	—	C
Latvian	1996	—	—	—	1996	—	C
Lithuanian	1993	—	1993	1991	—	1996F Lith/En	C
Macedonian	1986	—	1986	—	—	—	C
Norwegian	1965	—	—	—	1965	—	C
Polish	1997	1996	—	1997	—	—	E&C
Portuguese	1997	—	1997	—	—	—	C
Romanian	1997	—	1997	1995	—	—	E&C
Russian	2003*	2003* s	—	—	—	1994F, 2001C	E&C
Serbian	2004	—	2004	—	—	—	O
Slovakian	2000	—	—	1981	—	2000C (Sl./En.)	E&C
Slovenian	1995s	—	1973	1994-1995s	—	1991F (abr)	O
Spanish	2004	—	2000	2004	—	2000C	E&C
Swedish	2003	1983	—	1977-1981	—	2003W (Sw/Fi/Span)	E&C
Turkish	1925	—	—	1925	—	—	C
Ukrainian	1999	—	—	—	1999	—	E&C
Vietnamese	1984	—	—	1984	—	—	C

96 *= in print/preparation, s = supplement, C = CD-ROM, W = Web, F = database file.

97 Data was verified using: C = e-mail correspondence; O = OPACs, E&C = UDC Consortium.

98 Correspondence with Azerbaijan confirmed the existence of a translation (abridged edition in the 1960s. Reference was requested but not yet received at the time of this report.

99 This is a regular UDC MRF file update.

100 The existence of two special editions in Bahasa Indonesia was confirmed via e-mail and a reference was requested but not yet received at the time of this report.

Additional information was obtained through this research relating to the Arabic and Greek editions. Rigby is the only source mentioning an Arabic translation as in progress in 1981. This was verified through correspondence with librarians in Saudi Arabia, the United Arab Emirates, Morocco and Tunisia but their catalogues showed no record of UDC translations in Arabic. The only mention of any UDC translation in Arabic is "an outline by Shiniti" in Syria (Estanbouli, 1980: 416). Contact was established with the initiator of a translation project in Egypt that was known to the Consortium. The explanation obtained was that the project was planned and then abandoned in 1993 because of a lack of funds for the translation licence (Salem, 2004).

Similarly, some research was made into establishing the situation with a UDC translation into Greek. The UDC is used in 36 libraries in Greece (3.6% of all libraries) (Bokos, 2000). Correspondence with a person involved in teaching UDC in a library school revealed that no translation (published or unpublished) of the UDC is available in Greek. It seems that a project to start a translation was planned at some point by the National Documentation Centre in Greece but this was abandoned for unknown reasons (Skretas, 2004).

The research results indicate that, in spite of existing interest and the use of the UDC, many countries are not capable of covering the cost of a translation licence together with the costs of translation and publishing. Arabic-speaking countries such as Algeria, Morocco, and Tunisia are using UDC in French. Often the number of users seems insufficient to justify the cost of the translation and publishing of a new edition. This is the case in Armenia, Kazakhstan, Tajikistan, and Uzbekistan where UDC is used in special and academic libraries only and occasionally replaces the BBK (Soviet library classification). On the other hand, the lack of new editions and translations in countries where UDC is well-known and used (e.g. Scandinavian countries, Switzerland, and the former Soviet Union) can be explained by the fact that groups of countries are using and sharing editions not in their native language (English, French, Russian).

When it comes to type of edition, in the 1990s there were only two full editions (Russian, Polish), two special editions (Latvian, Ukrainian) but 13 medium and 14 abridged editions.[101] New trends in classification use are made evident through the release of 13 different electronic tools (4 web tools, 5 CD-ROMs, 4 database files) and it is noteworthy that only 4 of them are bilingual. The reasons are very different from those that impeded multilingual editions before the 1990s. While the MRF data are available in English, Spanish, Russian, Japanese, Czech, and Slovakian databases, a lack of multilingual electronic products is an interesting paradox in the competitive market of indexing tools.

This survey and the data collected indicate that the number of UDC translations is greater than previously estimated in the literature. But, the new changed ownership and management of the UDC, however, did not seem to have directly influenced or encouraged any new UDC translations.

7.2 UDC management, maintenance and distribution

Classification management built around the UDC system comprises: a policy for the development of the scheme, long and medium-term plans, revision and maintenance, publishing, marketing and promotion (Strachan, 1990: 4). In theory, each and every one of these work areas relies on the fact that classification schedules are stored in a database. The plethora of functions that are based on the database was more assumed than widely elaborated in the published literature on MRF in the period from 1993 to 2004.[102] The full exploitation of the database should justify its cost, its maintenance and further development. This research indicates that there is a certain amount of discrepancy between the potential of the UDC MRF database and its actual use in revision, or its use in classification marketing, selling, application building etc.

7.2.1 UDC database

In the opening paragraph of the report on the UDC MRF database creation project, Strachan & Oomes (1993: 19) cited the following Task Force recommendation as important guidelines behind their work: *"A standard version of c.*

---

[101] Since 1992 the following terminology for UDC editions is used: extended (instead of full edition), standard (instead of medium edition) and abridged.

[102] For the purpose of this research, apart from the Master Reference File Manual, a total of 8 articles on the MRF were consulted, 6 of which were written by G. Riesthuis (1997, 1998, 1998a, 2001, 2002, 2003) and 2 by Strachan & Oomes (1993, 1995). Brief information on the MRF can be found in Gilchrist's article in 1992 and in numerous subsequent articles written by I. C. McIlwaine (1993, 1995, 1997a, 1998a, 1998b, 2000b).

60,000 subdivisions, in English, in a machine-readable format should be created. It should be supported by a semantic network and have a much more consistently faceted structure than at present".

This statement illustrates an important concern of the then Task Force with respect to data content consistency, faceted structure and semantic linking. The term 'semantic network' gives the impression that the Task Force did not envisage the UDC database as a mere electronic text container, but as a possible basis for creating a thesaurus. The creation of a database was just one part of the MRF construction as described in detail by Strachan & Oomes (1993). The process went through an elaborate procedure of creating 30 subject sections based on tapes of the BSI Medium Edition 1985, their conversion to ISO 2709 file structure and their import to the previously created database, following the cancellation and input of extensions. These files were then merged to form an MRF of 60,000 records.

The choice of software is usually governed by the application requirements and the specific functions that the database has to satisfy, and this was not entirely the case with the automation of the UDC. It is fair to say that CDS/ISIS[103] was the only affordable solution for the Consortium, or to be more precise: the best of all affordable solutions. The strengths of the software could be summarised as the following:

- as information retrieval software CDS/ISIS offers a robust and reliable inverted file system.
- database design and management expertise can be acquired in a relatively short period of time
- efficient data storage
- easy modelling - all the related data are stored in one record
- supports an international standard for bibliographical data exchange (ISO 2709[104]
- CDS/ISIS is said to be the most widely used information retrieval

[103] CDS/ISIS is textual storage and retrieval software developed by UNESCO since 1985, mainly to satisfy the needs of developing countries for information storage using inexpensive technology.
[104] It should be noted, however, that ISO 2709 used in CDS/ISIS is slightly different from the same standard used in the bibliographic domain.

package in the world, especially in libraries (Buxton & Hopkinson, 1994)

- low implementation costs – it is supported by UNESCO CDS/ISIS and is distributed by national centres for a symbolic sum of money, or for free. These centres provide passwords for online download of upgrades and different CDS/ISIS software[105]

The requirements for the choice of database were neither published nor made available to a wider audience and the impression, from existing literature, was that the authors of the database, G. Riesthuis and D. Strachan, were aware that the system was not user-friendly nor easy to manage and that it had serious shortcomings in terms of data management and data protection (Riesthuis, 1998; Strachan, 1990). The release of CDS/ISIS for Windows (WinISIS) provided a much needed, improved interface and some other functions, but for these to be fully exploited some changes in the database format were necessary that were not taken into consideration by the Consortium. Also, the WinISIS application had costs in terms of the speed of free text searching, which is one of the most frequently used search functions in the maintenance of the UDC MRF. Hence, DOS CDS/ISIS continued to be used in parallel with WinISIS.

One of the primary rôles of the UDC database was the distribution and production of accurate UDC schedules. This was an easy task to solve with CDS/ISIS, which has powerful formatting and printing capabilities. Also, the way the UDC MRF database is designed allows for its use in different applications because of the relative control over data elements.

UDC classification data are represented in a way that data and structural characteristics are preserved and partially separated from the classification display and print layout. Each concept, i.e. classification number, is presented as a single record with information necessary to enable reconstruction of schedules in printed or

---

[105] Although the use of CDS/ISIS is widespread and most CDS/ISIS software can be downloaded for free, it is very important to note that CDS/ISIS is not open-source software. This presents a problem for bug fixing or software extension, and the availability of software documentation. It was reported that there were problems with additional libraries and information about thread safety and UNICODE compatibility. It seemed that it was not easy to build satisfactory web applications using a CDS/ISIS database. This is the reason why an initiative called OpenIsis was established in October 2002 with the goal of building an open source version of the CDS/ISIS software (http://www.openisis.org/openisis/doc/OverView). In June 2003 the first version of OpenIsis was released.

display format. In addition to this, every record contains other information relevant for revision and administration of the classification, as illustrated in the following section.

7.2.1.1 Database format

An MRF record consists basically of classification data and administrative data[106] (Master Reference File: manual, 2003). Further examination of the purpose of these data, however, reveal several functionally different groups of information held in the database.

Classification data are a set of 15 fields that hold data necessary to 'reconstruct' the content of classification schedules (Table 7.5).

Table 7.5: UDC MRF classification data fields

FIELD NAME	PURPOSE
UDC number	Primary data, UDC filing, Identification function
Table	UDC filing (database sort)
Special auxiliary type	UDC filing (database sort)
Combination type	UDC filing, management
Derived from parallel instr.	Links to the class on which the parallel division was made (number only)
Instruction for parallel inst.	[Divide as] UDC (number and annotation)
Parallel division provides special auxiliaries	If spec. aux. are product of the parallel division this states which type of spec. aux. is produced
Application note provides special auxiliaries	The statement of the type of auxiliary that should be excluded from parallel division
Description	Caption (UDC number description)
Verbal examples	Examples of the core concepts i.e. Including:....[107]
Scope note	Scope note - semantic explanation
Application note	Technical instruction for the application
Examples of combination	Contains examples of combination and their caption

Publishing/editorial data are two fields in the MRF record that hold data necessary for publishing or printing. The database administrator provides a textual description for each special character that appears in the record so that the original character can be traced irrespective of the changes that occur when the file is opened

[106] Full details on the MRF data structure can be seen in Appendix 6, Table 6.1.
[107] Example of verbal example following the caption: *=923 Systems derived from Volapük or Esperanto.* ***Including:*** *Ido.*

using a different character code table.[108]

Also, there is a field for editorial notes. This field is used when there is a need to provide information on a correction in the automatic filing of UDC or to store notes or comments that may be necessary for publishing or printing, but where this information cannot be introduced in any other field of classification data (Table 7.6).

Table 7.6: MRF data fields important for publishing

FIELD NAME	PURPOSE
Use special characters	UDC publishing, display
Editorial annotation	UDC publishing, display

Revision administration data (Table 7.7) are a set of 16 fields that are populated in the process of a database update as part of the revision process and MRF administration. Five of these fields are necessary for basic MRF maintenance and UDC distribution and production. The remaining fields are envisaged for supporting UDC revision management (planning, supervision and control). These fields are for internal use and are not included in the user version of the MRF (UMRF). A few fields have to be distributed in the cancellation files (cancellation date, replaced by) otherwise users are not aware of the existence of these data.

Table 7.7: MRF administrative data fields

FIELD NAME	PURPOSE
Introduction date	Revision control and management
Introduction source	Revision control and management
Introduction comments	Revision control and management
Cancellation date	Export and deletion of the cancelled numbers, production of MRF CAN
Replaced by UDC	Export and deletion of the cancelled numbers,
Cancellation source	Revision control and management
Cancellation comments	Revision control and management
Last revision date	MRF maintenance
Revision field	Revision control and management
Revision source	Revision control and management
Revision comments	Revision control and management
Revision history	Revision control and management

[108] This is in a way related to the limitation of CDS/ISIS in representing characters outside the extended ASCII table (Master Reference File, 2003).

FIELD NAME	PURPOSE
Index only UDC notations	Revision control and management
Admin notes	Revision control and management
For next E&C	Revision control and management
Temp work only	Revision control and management

Record management data (Table 7.8) hold information on record identification and the database that is a part of the CDS/ISIS database management program. The software does not allow full automation of data management and many functions for MRF maintenance are based on manual handling, validation and tracking record numbers against the database name. A record number is the only UDC number identifier that is independent of its notation and reflects its relative position in the database. The MRF Manual provides instructions for database naming and this is an important part of MRF maintenance and distribution.

Table 7.8: MRF record management data

DATA	PURPOSE
database name	identifying MRF database it belongs i.e. MRF, UMRF, CAN (MOD, NEW), this information is independent of the exported file name
record number	MFN (master file number) is a record number automatically assigned by the program

Appendix 6.1, Table 6.1 shows statistics for the population of the UDC MRF 2003, which was valid until March 2005 when the next version, MRF 2004, was distributed. The search performed against fields to produce these statistics, i.e. presence and absence of data in a certain field, is part of regular MRF maintenance. Such statistics often indicate some kind of irregularity that needs attention from both the MRF maintenance body and revision committee as is indicated in the comments against such situations in the table.

7.2.1.2 Database functions[109]

The UDC MRF database and its maintenance are created around and made fully dependent of the standard CDS/ISIS functions (DOS). No upgrades were made

[109]The parallel reference to the DOS and WINDOWS version of CDS/ISIS is necessary because both have to be used in UDC MRF maintenance. While WinISIS is used for editing and export, as it provides a better interface, the DOS version is used for free text searching and data print/export/import which is too slow in WinISIS.

to the original format to make better use of WinISIS, and there were no utility programs or new data export formats in the period 1993-2005.[110]

Import/export of data are based on manual record manipulations and the import/export function is one of the most important in the MRF update and distribution process. This function has several specific purposes:

- addition of new records (import: merge/load)
- creation of distribution files (export: through different FST field selection table)
- sorting of the database (import of the database exported through print/sort function)
- removing empty records from the database after deletion of cancelled numbers (whole database export/import)

Record numbers, called *master file numbers* (MFN) in CDS/ISIS, are added automatically by the program. Each record represents one UDC number (notation) but the relationship of the UDC number to its record number changes in the process of an MRF update. When numbers are cancelled they are exported from the MRF, the database is compressed and all records are renumbered. When new numbers are added the whole MRF is exported through a print-sorting format and imported into the database again. This is performed to retrieve the correct sequence of UDC numbers. As a consequence, MRF record numbers cannot be used as permanent identifiers of the content they represent. This is tedious if end-users want to use them for the automatic update of their UDC schedules.

Editing is probably the weakest part of CDS/ISIS, and especially the DOS version. Data entry in CDS/ISIS requires that field and sub-field tags be keyed in, which slows down the process and leaves room for mistakes. But more importantly, there is no validation or control over field content and DOS CDS/ISIS, for instance, allows for field tags to be omitted or mixed, without warning, or it allows for duplication of UDC numbers. Lacking automatic control, both field content and field

[110] The only exception being the UDC MRF Scanner program (see Appendix 6.2) which was created on a voluntary basis (not at the request of the Consortium), to help the database administrator in coding the database import. The program and Manual were then handed to the Director of the UDC Consortium in 2000 and were updated in 2001.

tags have to be checked manually with the help of free text searching and printing of control lists. DOS CDS/ISIS is limited from the data management point of view as it allows an empty or invalid record to be saved. Editing can be improved when performed in WinISIS, which allows copy/pasting from other Window applications, better management of special characters and easy selection of fields to be edited.[111] The only positive side of the CDS/ISIS editing function is the easy transition from search to editing, which allows for whole search result sets to be opened for editing.

Display/Print is a good feature of the program. Using the CDS/ISIS formatting language, it is possible to create an unlimited number of display/print formats. The display/print format contains the choice and presentation of the fields to be displayed or printed and once created display/print formats can be use to print, sort or export the database file. The rules for display and print can be written 'on the fly' using a formatting language which is often used for cross checking. The default display/print format in UDC MRF is MRF.PFT and when used it shows and prints the content of the complete record together with the field/sub-field tags of all fields. The UDC MRF is distributed to users in a 'user format' (UMRF.PFT), which excludes the administrative and maintenance fields of the UDC MRF. Other formats created and available since 1993 are EXPORT.PFT, EXPCAN.PFT, EXPMOD.PFT, TABLE.PFT, and TABLES.PFT (MRF Manual, 2003). These print formats are one of the strengths of CDS/ISIS and more formats catering for different user needs would have been easy and inexpensive to produce.

Browsing is very poorly supported by CDS/ISIS. The only browsing function supported by CDS/ISIS is the browsing of database records numbers (MFN). To do so, one has to enter the range of records to be browsed. This means that browsing the classification hierarchy is only possible if it follows the sequence of database records. But even if this is the case, one has to know the exact record number from which to start browsing in order to see the part of the schedules of interest. One can also browse up and down throughout the term dictionary and by selecting a term, open a record. WinISIS improved browsing to a significant extent by allowing two

---

[111] WinISIS allows for other advantages in editing to be added such as, 'pick up' list or mandatory fields, but this would require the adaptation of the existing database format and there is no evidence that this was considered an option by the Consortium.

windows to be scrolled: one browsing the record display and the other browsing the term dictionary holding UDC numbers. But to see where to start browsing one has first to perform a search of UDC numbers, see the number of a record the UDC is occupying and then go to the browsing window to type in the record number from which to start. This limitation of CDS/ISIS is the reason why UDC numbers in the database have to correspond to the sequence of the MFN and why the database has to be re-sorted every time records are added or deleted.

Sorting in the database is usually controlled on two levels: data sorting and record sorting. Sorting of data in an CDS/ISIS program is defined by the ASCII character code table which causes wrong ordering of UDC numbers in the MRF as the symbols used in UDC require a different ordering sequence. The MRF Manual has instructions for managing this and shows how to adapt the file that controls character conversion: ISISUC.TAB (this table is distributed to users as a file called UDCSORT.TAB).

The sorting of UDC MRF records in order for them to follow the proper sequence of UDC numbers is not an automatic procedure. The whole MRF database has to be sorted after each update and this is achieved through a print/sort function. The whole database is sent to print file through a sort form that defines two sort keys: the first sort key is the table field (a code for the classification table from which the UDC number originates) and the second key is the UDC number itself. The sorted database file is then imported back into the UDC MRF.

Searching, of all database functions, is the best supported and one real advantage of having the UDC MRF in a CDS/ISIS database. Searching is based on a robust and efficient permuted file indexing technique. The technique of field indexing is completely open to change and alteration. Searching in CDS/ISIS supports right truncation, phrase searches and Boolean logic in a very powerful way. Every search result can be saved as a hit file and then used for printing, editing or export. The search can be performed through a term dictionary that lists all the terms indexed by the database. But most importantly for MRF maintenance is a free text searching function that allows search parameters to be launched against specific database field(s). This search is not based on an inverted file but rather on the full content of the database and is comparatively slower. In WinISIS the function of *free text searching* is so slow that, when it comes to the UDC MRF maintenance, it is impossible to work with.

7.3 Database in supporting UDC update

The annual MRF update comprises not only the manual editing of UDC number modifications but also systematic cross-checking of data on all levels, as CDS/ISIS is not a database management system and does not have provisions for data validation, protection or tracing. This means that UDC number strings can be duplicated or even missing, references can refer to non-existent classes, examples of combination can contain non-existent UDC numbers or cancelled numbers. Checking can be done only by a visual field-by-field free text search, looking for the presence or absence of data and counting records. The database administrator has only enough time to do basic checks for duplication of the main number field and whether the total number of cancellations, new numbers and cancelled numbers correlate.

Because of the fact that the output of the revision work prior to its publishing in E&C is not double checked using a database tool in order to crosscheck the notation and concept occurrence across the database fields, the output may contain incorrect notation.[112] In the process of the database update, data may be deleted accidentally or by following incorrect update instructions. Once the update is completed, no checking of the database file is performed by the revision committee prior to UDC MRF distribution.

Import of new numbers into the MRF is the most demanding and time-consuming of maintenance jobs in the UDC update. To add even more to the total number of update tasks, the list of new UDC numbers to be entered in the database is provided as an unstructured MS Word document simply formatted for print in the *Extensions & Corrections to the UDC (E&C)*. This document is missing a significant amount of information necessary for updating the database.[113] The database administrator has to code, and in some cases additionally investigate, the following: the source notation to be indexed in the case of pre-synthesised notations, the source

[112] This issue will be explained in more detail in Chapter Eight in the section on data validation. Authors of the schedules use only the print output file from the database based on the user export format, which contains only a minimum of database fields.

[113] Not up to the 2004 update was the revision output changed to introduce at least the indicator of the type of note (AN for scope note, SN for scope note and EN for editorial note). This will help the database administrator stream these data into appropriate fields automatically without the danger of misinterpreting the original intention of the schedule authors.

and target notation in parallel divisions and also what data are supposed to be application notes, scope notes or editorial notes.

From this file, which may contain two to three thousand numbers, a plain text file of numbers to be added needs to be extracted.[114] After this is done it is necessary to produce a structured and encoded text file. Shown below is an excerpt from *E&C 21*:

Example of the file format that is delivered for MRF update:

+004.455.1	Client software. Including: Browsers *Example(s) of combination:* 004.455.1:004.738.52 World Wide Web browsers
+ 141.78	Postmodernism ⇒ 7.038.6
+159.944.4	Fatigue. Stress
! 165.611	Intuitionism
- 331.441	Strenuous, physically demanding work. Including: handling of heavy goods
x331.444	(Influence of length of work period) → 331.443:331.31

From this the choice of numbers that needs to be imported is made:

+004.455.1	Client software. Including: Browsers *Example(s) of combination:* 004.455.1:004.738.52 World Wide Web browsers
+141.78	Postmodernism ⇒ 7.038.6
+159.944.4	Fatigue. Stress

A plain text encoding (the input file) for conversion to an ISO2709 file of the three UDC numbers from the example above:

```
$$<CR><LF>
001 004.455.1<CR><LF>
100 ^eClient software<CR><LF>
105 ^eBrowsers<CR><LF>
115 ^b004.455.1:004.738.52^dWorld Wide Web browsers<CR><LF>
$$<CR><LF>
001 141.78<CR><LF>
100 ^ePostmodernism<CR><LF>
125 ^a7.038.6<CR><LF>
$$<CR><LF>
001 159.944.4<CR><LF>
100 ^eFatigue. Stress<CR><LF>
```

This encoding can be produced manually, using macros. But for several thousand

[114] Numbers that need to be added to the database are marked with +. UDC numbers that have some text modification are marked with ! and those that remain the same are marked -. Numbers that need to be cancelled are marked x. In the E&C they are all filed together. For MRF maintenance they have to be separated.

records that may need to be added to the MRF this can take up to a month. In the year 2000 a computer program application called *MRF Scanner* was written to speed up and control the encoding (Appendix 6.3). The program provides most of the encoding automatically and allows easy checking, editing and correction of the output, which is then exported as a file ready for ISO 2709 conversion. Fields such as *Introduction data* and *Introduction source* are automatically added. This reduces the input preparation that would otherwise take a significant amount of time. *MRF Scanner* can be further developed to provide full encoding, but its limitation lies in the poor structure of the original file from the *E&C*.

Once this output file (see example above) is produced, based on the instructions from the MRF Manual, the conversion to an ISO 2709 file can be done using a program called FANGORN2. Apart from the input file, FANGORN2 needs a specification file that can be produced using the MRF Manual. Once this has been done, FANGORN2 is run on the encoded file and the product is an ISO2709 file of the records that can be imported into the CDS/ISIS database.

This is an example of the ISO2709 file format produce by the program for the first of three UDC numbers in the example above:


```
00199000000000109000450090300060000000200020000610000180000810500110002611500370
00379010005000740010010000079#E&C21#M#^eClient software#^eBrowsers#^b004.738.52^d
Worldwide Web browsers#9912#004.455.1##
```


Once imported into the database, there may be other data to be entered such as a note in the *field of use of special characters* - field **952**, which is a consequence of CDS/ISIS not being capable of handling an extended ASCII set. Another field to be completed is the *index only UDC notation* field **951** which lists all the occurrences of numbers in the examples of combination, application and scope notes, as they are entered as a simple text string and cannot be indexed. Every UDC number needs to be found and accessed when the original number is cancelled. The process of supplying index notations if performed thoroughly would significantly increase the time of the UDC MRF update as the database administrator has to trace back thousands of pre-built notational elements and has to edit each of them in the database.

Modification and cancellations - cancelled numbers are first manually edited one by one to fill in the cancellation date, replacing the UDC notation (if any), source of cancellation (E&C) and comments (if any). A search is then made for all edited records (free text search for the date of cancellation) and the results are exported as ISO files and through print formats to produce text files (called CANyy, where 'yy' represent the year of update). The marked records are then deleted from the UDC MRF following the free text search for *all records marked for deletion with a current date*.

Cancellation of numbers will trigger many modifications of the remaining records, as all the occurrences of the cancelled numbers in the examples of combination, references or notes should also be removed. In order to be sure that the cancelled number is removed/replaced from all MRF records, each number is searched individually. A search will not find matches if these numbers are used in examples of combination and pre-synthesised notation elsewhere or if each of source notation is not entered in the *index only UDC notation* (field 951).

The number of cancellations adds to other modifications notified by the revision committee. The same procedure of export as an ISO file and printing as a text file is repeated for *all records marked as modified with a current date*. This is how MODyy files are created for distribution.

The whole MRF update process seems to be reversed from a management point of view.[115] A reasonable course would be to produce a properly structured text file at the end of the revision process with all the necessary instructions. The process of entering these changes to the database would trigger consistency checking at a very early stage and once they are in the database, numbers are accessible for searching and further checking. Once this is done an output file for publishing in *E&C* can be produced in a straightforward manner and this can be easily double-checked before publication for any additional changes. In this way the content of the UDC MRF would be checked twice, once when entering data in the database and the second time while preparing the *E&C*, which would give enough time for additional corrections before UDC MRF distribution. The advantage of this approach would be

[115] The following comment by the creators of the UDC MRF indicate that the original plan was that database update is starting point for the E&C: "The database should be completed within two years, and provide the individual publishers of UDC versions the material for the compilation of their editions. It should also be the basis for revision of the schedules and the **starting point for** Extensions and Corrections to the UDC" (Strachan & Oomes, 1993: 19)

that the UDC update and production of distribution files would take significantly less time, while the quality of the final output is likely to be better. This approach, however, requires more management effort, strict time management and revision output in the form of richer and better structured textual instructions.

The way in which the revision output is organized indicates that the UDC maintenance management is based on one of the following incorrect assumptions: *(a)* it was assumed that the database administrator is very knowledgeable about the UDC and has plenty of time to split and check manually every pre-synthesised notation submitted in an update, which may amount to several thousand records, or *(b)* it is assumed that the UDC MRF database tool is an advanced intelligent tool able to perform automatically the task of tracing and automatically decomposing UDC numbers in whatever form they may be submitted for update.

The problems that arise from the present management and the maintenance of the UDC can be attributed to a combination of factors:

- the nature of software (i.e. limitation of CDS/ISIS in data management) which asks for more manpower, better planning and strictly controlled management of the maintenance
- a lack of resources or/and expertise for the Consortium to plan, provide and fund additional tools and human resources that would make up for the shortcomings of the existing database tool
- a lack of understanding of database management and data quality issues on the part of the Consortium, which results in a low priority being given to database use in the revision process, maintenance and support
- the poor management and lack of coordination between revision work and update which results in the revision output being formatted for printed publications and not for database update. As a consequence the changes published in *E&C* are checked twice but no editorial control or checking is established over data entered in the MRF database

Because of the fact that maintenance of the UDC MRF is essentially a manual and time-consuming process, it is up to the UDC owner to ensure that this does not affect the quality of the final product.

7.3.1.1 Different UDC MRF 'databases'

UDC updates are managed by keeping data in several different "variants" of database files, which are stored and distributed as their annual editions. First of all, there are two editions containing a complete set of approved UDC numbers: MRF and UMRF. Each year there are also exports from the database that can again be kept as separate databases: annual UDC cancellations (CANyy.ISO), annual extensions (NEWyy.ISO), and annual modifications (MODyy.ISO).

MRF is the main database file i.e. the actual Master Reference File and it contains the full data of the last updated version of the UDC. This database is used for maintenance of the current schedules. Each record contains a complete set of fields. This database file is not normally distributed to users.

UMRF (i.e. 'users' MRF) is the distribution version of the MRF. After the MRF is updated it is exported into the UMRF and named after the year e.g. 0012 (yymm - 2000, December). These versions exist for each year starting with 1993, and all annual versions are kept for historical reference. What is distributed to the users from the actual MRF is text (ASCII) or an ISO 2709 export of the data without MRF administrative fields. The text export (UMRF.txt) contains a selection of classification data sufficient to print UDC schedules only. It lacks the following data: a table code, (describing the table from which the number comes), special auxiliary type code, text language code and all administrative data. The ISO 2709 export, on the other hand, is simply an MRF UDC database export (minus administrative fields) as it contains database specific sub-field tags. This export can be used for recreating the UDC database in CDS/ISIS or alternatively for converting into some other more convenient database structure file. Separate files with records that were added, cancelled or modified are distributed to users together with an annual edition of the UMRF (also in text or ISO format).

MRFCAN is a file of cancelled numbers that could be merged into a single database to hold information on deleted and removed database records. Once they are deleted from the MRF, cancelled numbers could only be tracked through the cancellation export file. Unfortunately, the database itself has never been created or maintained although the files for its creation are regularly produced. Before they are deleted from the UDC, cancelled number records have the following fields populated: *date of cancellation*, *source of cancellation* (e.g. volume of *Extensions & Corrections to the UDC* in which cancellation is published), and *replaced by*. This information is not held reciprocally in the records of the new numbers that are introduced as a replacement.[116] The cancelled UDC numbers are exported as ISO and text files and then deleted from the database. The amount of cancelled records per year is in the range of a few hundred to a thousand.[117]

MODIFIED RECORDS ('MRFMOD') stands for the annual file containing records that have had some alteration in the text of the record. Corrections of this kind are done manually in the database, record by record and each change is noted in the revision fields (*revision date*, *revision source*, *revision fields*, *revision comment*, *revision history*). Modifications for each record are marked with the year of change and specification of changes made. Once the modification is finished these records are retrieved and exported as text (MOD*YY*.txt) or an ISO2709 file (MOD*YY*.iso) and distributed to the users annually together with separate files containing new numbers and cancelled numbers. Usually several hundred are modified each year. If merged with the 1993-2004 modifications, a 'MRFMOD' database would be useful for tracking and monitoring common mistakes and building a policy for validation and checking of the MRF.

NEW RECORDS ('MRFNEW') is a file containing the annual addition of new records.[118] Records are not edited directly in the database, but through a process of manual text encoding, ISO2709 conversion and database import (see previous section on Database functions - import of a new records). Each newly added record is traced through the following administrative fields: date of introduction, introduction source and introduction comment. Once records are merged into the database and sorted, so that they fall into their correct place in the UDC, they are exported as text (NEWyy.txt) and a database file (NEWyy.iso) and distributed to users. It is possible to create a database (MRFNEW) of 'new records' that may be used if one wishes to produce some statistical analysis of revision policy, the total amount of numbers added, the distribution of new numbers across the UDC etc.

---

[116] This information would be very relevant for management of the revision of the UDC, but even more so for bibliographic services using UDC over a period of time.

[117] Since 2001, text files have been made available for download at the Consortium website (http://www.udcc.org/cancellations.htm).

[118] Statistics for the field *901 Introduction Date* in Appendix 6.1, Table 6.1 show the number of new records introduced for each year between 1993-2003, the largest being over four thousand in the 2001 update.

It would make sense if these annual 'historical' data files, MRFCAN, MRFMOD and MRFNEW were held in a single UDC **master** database holding together all UDC data. This could then be used for supplying 'on-demand data extracts' for changes, new numbers or cancellations in a certain subject area or in a certain period of time. Producing these data as individual files and distributing them as annual updates does not make much sense. Users buying the UDC MRF, for instance, need cumulative cancellation data that cover the entire period prior to their buying their first UDC MRF. In other words, it would make sense for them to buy a complete UDC MRF not just the latest edition. Also, they may need to continue purchasing cumulative modifications, cancellations and additions once their licence expires. None of these is made available, offered or marketed to users at the present, in spite of the fact that these data already exist and require no additional effort or expenditure by the Consortium.

It is unusual in database management that records with cancelled numbers are actually physically removed from the database, especially as UDC cancelled numbers remain valid in practice and users need to maintain compatibility between new and old UDC notation. The Consortium's approach to management of UDC is not entirely clear and no documentation was found explaining the present 'packaging' and distribution of UDC data.

7.3.2 Some issues in UDC MRF distribution and application

CDS/ISIS database files necessary for recreating the database that holds the MRF (Aumrf.fmt, Bumrf.fmt, Table.pft, Udcsort.tab, UMRF.fdt, UMRF.fst, UMRF.pft, UMRF.stw) are distributed free to users together with the MRF Manual. Using these files and the usual CDS/ISIS procedures, it is easy for each user to create the sets for his/her own databases to manage the UDC. However, in terms of information exchange and interoperability, the MRF database is only an intermediary and separate proprietary tool that cannot be connected to a library system.

A three-year licence governs the purchase of the three subsequent annual editions of the UDC MRF.[119] This is, obviously, provided on the assumption that users (publishers and cataloguing agencies) would welcome each new version of the

[119] Categorisation of UDC MRF licences is discussed in more detail in Appendix 6.4.

UDC MRF that they will somehow merge/load with their IR system, classification authority file, translation or classification tool. The fact is that the data in UDC MRF annual distribution files are not backwardly compatible with previous versions and cannot be used to update automatically UDC data already held in a local system. The reasons for this are: a lack of a permanent UDC number identifier, the deletion of records in the UDC MRF, the re-initialisation of UDC record numbers in the MRF database and the re-use of cancelled UDC numbers.

Because of the nature of UDC MRF management, data format and the poor choice of export formats, the use of UDC MRF is coupled with significant costs. In this context the purchase of a licence on an annual basis is in contradiction with the real value of the product for buyers. A library would only be able to update their existing database annually if the process were automatic and, at the present, the UMRF data do not satisfy minimal technical requirements for this process to be automated. From a technical point of view, every new version of the MRF is not a simple 'update' for the library but rather an entirely new UDC database to be re-implemented in a library system. Also, the idea that libraries can use an annual edition of the MRF to update their implemented version of the UDC cannot be related to any existing library practice for the following reasons:

- the ISO2709 export of the MRF is not available in any of the MARC formats widely used in libraries and as such is incompatible with a library system
- provided that it is available in one or several MARC formats, a UDC MRF has to be built around permanent UDC number identifiers

The distribution of the UDC is still driven by producing the UDC for publishing and tends to neglect the way UDC is and is likely to be used. UDC MRF distribution management also indicates that the Consortium may be basing its product on the assumption that the UDC is not used in online systems at all, and that the only purpose of UDC MRF is to be a standalone 'reference tool' consulted in the process of classification in the same way as was the case with a printed edition. This clearly does not correspond with the reality and the UDC MRF is often used as a basis for classification tools. It would be commendable to make its future distribution more adapted for this particular purpose.

Although anyone can purchase and use the UDC MRF database for its application in bibliographic services, this database is not designed for use by classifiers. In addition, the classification browsing capacity and interface, in general, are not the strongest side of CDS/ISIS, which makes the original UDC MRF database a poor candidate for an end-user application. This is the reason why the UDC MRF is likely to be used as the basis for producing user-friendly classification tools and this has been, up to now, exploited by several publishers.

The first series of these tools are the 'UDC schedules on CD-ROM' applications. In 1997 the first bilingual English-Czech version of the UDC MRF was published on CD-ROM and was followed by an update in 2000 (Appendix 7.1.2). A Spanish CD-ROM was published in 2000 (Appendix 7.1.1). Slovak-English and Japanese-English CD-ROMs followed in 2002. A Russian tool on CD-ROM was published in 2001 with an English-Russian subject-alphabetical index. All these electronic versions have a user-friendly browsing and searching interface.

The greatest breakthrough for end-user tools was the introduction of schedules on the Web. The best examples of these tools are BSI's *UDC-online*[120] (http://www.udc-online.com) and the Czech Mezinárodní desetinné třídění - UDC database (http://aip.nkp.cz/mdt), which have been available since 2001 (Appendix 7.2.1 and 7.2.2). Access to BSI is regulated through an annual licence while MDT can be accessed for free. The UDC on the Web has great advantages for users both in the flexibility of access and in speed of update. The BSI product has an especially well developed user-friendly searching and browsing interface, an easy transition from search to browse, and intuitive navigation and also contains a number-building function.

The Swedish Universella Decimalklassifikationen - Svenska elektroniska utgåva (http://www.hb.se/bhs/udk/) is a selection of around 7000 UDC terms available for browsing online for free (Appendix 7.1.3). Although this is not comparable with tools developed by BSI and the Czech National Library that are based on the entire content of the MRF, the Swedish schedules are an important

[120] This BSI product was created in cooperation with UK software company Technical Indexes (TI), which was responsible for technical support. At the end of 2004 BSI re-created this product with the same interface functionalities as the first product (Appendix 7.2.1).

reference source for Swedish libraries using UDC (Benito, 2001).[121]

All of these tools are based on UDC data held in relational databases and the main cost of each of these tools is to create a database and convert the UDC MRF to data that can be used in supporting browsing and hyperlinking. No existing UDC MRF export is suited for this particular use. A text export, for instance, does not contain the data necessary for managing tables. MRF/UMRF data, on the other hand, contain structure and tagging used in CDS/ISIS that are completely non-transparent and unnecessarily complicated for the creation of a new database.

It is obvious that different new and improved exports of UDC MRF are necessary for those who publish and use UDC electronically. A far better quality of UDC MRF output can be achieved even with the existing database structure. On the other hand, a new improved version of the UDC MRF database format is paramount in meeting the needs of libraries and those that use UDC in IR. The proposal for a new UDC MRF format and the requirements for the improvement of UDC source data will be explained in more detail in Chapter Eight.

7.4 The use of database in UDC revision and maintenance

The main task of revision is to keep the content of the UDC up to date with knowledge growth so the system can be continually used for knowledge organization of both historical and contemporary information resources. It could be suggested that, from the 1990s onwards, the UDC has been used predominantly in an online environment. It is, therefore, a reviser's task to make sure that the UDC is fully functional in information retrieval and that unpredictable structural or notational instances are minimalised as they are the most serious obstacles in classification automation.

UDC content revision and maintenance need to ensure a consistent level of specificity throughout the system. At the same time, the UDC is designed for many purposes and its development should be ruled by literary warrant rather than by its

[121] Similar to the Swedish online edition is *La CDU Online*, which is an Italian web application of abridged UDC schedules created for use alongside the catalogue of the Archivio Collettivo Nazionale dei Periodici (Appendix 7.2.4). It is not based on the UDC MRF but rather on the abridged Italian translations of the UDC from 1987 (http://mail.biocfarm.unibo.it/%7Espinelli/cdu/).

potential application and implementation. This means that, for instance, UDC content as it is held in the UDC MRF, is not supposed to be orientated towards a 'helpful shelf order', 'the need of libraries', or the 'needs of the World Wide Web', although all these applications may be taken into consideration. Therefore, specificity level, application instructions, scope notes, and comments should not be formulated assuming or suggesting any particular use. The user study at the beginning of this chapter shows that UDC is still used worldwide, which means that its revision should not be biased towards its use in any specific geographic or cultural environment.

The original design of the UDC is based on the principle that makes automation easy: permanency of relationships between schedule content, notational representation (notation) and syntax functionality. Consistent syntax rules are set at system level as well as at individual class level. The content of the classification and the organization of concepts into facets and hierarchies are made functional through appropriate data formalisation and representation summarised in notation. Accurate and consistent notation is a basis for the UDC analytico-synthetic syntax, which is created for the purpose of indexing/information retrieval. It is a revision task to make sure that the classification notation always underpins the function for which it is meant:

- express hierarchy
- indicate the facet it belongs to
- indicate the table it comes from
- indicate the type of number: single, compound, complex
- indicate the combination of elements if the number is compound

The revision task is, also, to ensure that expansion of the system does not violate these rules and that the functionality of the system does not jeopardize expansion of the schedules. This is not always easy and at any given time in its history, the UDC has contained a number of faults on the level of content, structure and classification data. This is the reason why the editorial work usually results in cancellations, and corrections (modifications), as well as expansion of the schedules. Each revision helps correct mistakes but also allows new mistakes to be made and precise guidelines and the training of people performing the revision are paramount.

The automation of classification is supposed to minimize unwanted mistakes or structural incoherence and imbalance on the level of classification content, structure, hierarchical and facet organization and syndetic linking. In the revision process, a database can serve not only as a tool for finding mistakes but also for marking them so they can be automatically listed for correction at any given time and held together with the UDC data. This would be in accordance with responsible system management, documentation preservation and would enable long-term planning for system maintenance.

The use of the UDC MRF database search and browsing in the revision process helps to discover, control and mark mistakes and it is regrettable that the database was not, actually, used for this purpose in 1993-2004. The field in the administrative data reserved for documentation of the faults and listing them for revision is empty. The relationship between the revision procedure and data quality will be examined in more detail in Chapter Eight when reporting on UDC MRF 2003 validation.

7.4.1 Relationships between classification automation and revision

The literature discussing UDC revision indicates that it was not always clear which issues in classification revision were supposed to be solved by automation and which by managerial decisions (compare for instance Lloyd, 1972, Newcombe, 1972, McIlwaine, 1990). It appears that sometimes revision decisions and developments were postponed or not made because it was expected that the automation of the schedules was going to alleviate the problem. This is, for instance, the case with the UDC management decision to commission a database without a field structure, which would support a subject alphabetical index, on the grounds that this is a matter for each publisher to sort out. So, not only is the cost of creating a subject-alphabetical index conveniently transferred to publishers, but publishers are also unable to use the existing database format for this purpose. At the same time, the same UDC MRF product is sold to users who also have to bear the cost of creating a proper search index to classification which makes the building of electronic editions and tools based on UDC MRF expensive, and time-consuming. These hidden costs are a deterrent for new and potential users. But what is even more serious is that the lack of an alphabetical subject index influences the way new captions, scope notes, and application notes are expressed as the choice of terms and synonyms have to influence by their suitability for searching.[122]

Some revision decisions such as subject relocation, cancellations or the introduction of new concepts and re-use of existing numbers to denote new concepts were accelerated in 1993-2004 by the fact that these changes are now quicker to implement and reach UDC users much faster. It is assumed that reclassification is easier and faster in an automated library environment.[123] This would be true if the mapping between older and new UDC editions were provided in the data format that would enable this to be automated, and if the changes were well documented, published and made easily accessible and manageable for users of the UDC, which is not the case at present.

When buying a UDC MRF licence one receives existing data and changes from the previous year to the current year; there is no UDC product package with backward compatibility built in that would allow the automatic tracking back of changes, cancellations and replacements. New UDC numbers do not contain data stating which old numbers they replace (if this is the case). There is no connection between the library system and the way they use classification and the proprietary UDC MRF distribution files, and no procedures, documentation and tools have been developed to bridge the two.

Occasionally, suggestions are made to remove some features of the UDC for the sake of automation. In the course of technological development, however, these suggestions kept changing as was explained in Chapter One. For instance, since the very beginning of library automation there were suggestions to 'simplify' UDC notation for the purpose of computer handling or to get rid of some synthetic functionalities under the pretext that they are not easy to understand and use such as range notation or notation based on parallel division. At present, it is clear that there is room for simplification and reductions of unnecessary variants in notation, in the area of special auxiliaries. Strict policy and guidelines need to be established and

---

[122] Apparently, according to the Editor in Chief of the UDC, a subject-alphabetical index to the UDC has been considered and rejected by the Consortium.

[123] It would be fair to say that, in spite of automation or maybe because of it, libraries have significantly fewer staff and more users and therefore few or no resources to handle reclassification.

followed with respect to the rôle of special auxiliaries in representation of knowledge structure across disciplines. The range notations (e.g. 592/599) and notation based on parallel divisions (e.g. 821.111 developed by adding =111 to 821), recognised as problems in automation, are unlikely to be completely removed from the UDC. Because of this, the problem should be addressed through additional data coding that would enhance automation and facilitate notation indexing and searching.[124]

There are, however, situations when the UDC was guided and influenced by the reality of the classification being used in an automated environment. One good example of revision changes triggered by automation is the removal of a 'final digit' from the special auxiliaries.[125] Because the 'final digit' did not provide a fully defined notation linked to a clearly defined meaning, this notational device was in contradiction to the basic formality necessary in automation. Also, revision work in the 1990s respected the importance of facet indicators in the automation of an analytico-synthetic classification. The relationship between revision and automation will be more specifically addressed in Chapter Eight, which will be looking into how changes introduced by revision influence the manageability of UDC data in the present database format.

7.5 Summary of UDC use and management

Research reported in this chapter indicates that the UDC is still the second most widely used classification system in the world. Used in at least 112 countries and translated into at least 39 languages, this system has a significant global user community. The UDC is, however, only one classification system in a competitive market and its future is likely to depend on the availability and affordability of classification tools, especially those that are multilingual. This research indicates that

---

[124] If the data are properly structured and coded it is possible to use algorithms for automatic indexing and searching of range notations and notation based on parallel division. Appendix 10.4 illustrates how the search for range notations is automatically solved in a relational table in the *UDC Research Database*. Parallel derivation, for instance, requires definition of source data and target data, which is already done in the UDC MRF. From this it would be possible to define a base number, its prefix and suffix, and derive algorithms for automatic notation generation.

[125] Special auxiliaries known as 'final digits', represented as ***...1/...9***, are not used in the form in which they are listed. When these numbers are applied, the three points are replaced and a final digit is merged with the main number without a facet indicator. This means that numbers created with a final digit cannot be controlled or searched.

there is a certain discrepancy between the potential of the UDC MRF database and its actual use in revision, or its use in classification marketing, selling, online application building, etc.

The database used to maintain and distribute the UDC is created using software with a low implementation cost. CDS/ISIS has known limitations in the area of data management and validation, in handling special characters and in supporting browsing. The maintenance of the database is to a great extent a manual process, which means that it is more expensive, time-consuming and risky in terms of data loss and corruption. When used, such a high-risk software solution requires additional time for updates, management effort, documentation, tools and database structure enhancements that can work as a safeguard and would ensure the quality of data output. But, this does not present a problem if CDS/ISIS underpins a business that is managed by specialists who understand the software, the nature and use of the final product and the market needs.

In the period 1993-2004 no resources were put in to the improvement of database maintenance, management or data formats and exports. The explanation given by the Consortium is that there was a lack of resources, primarily financial, in this twelve-year period. As illustrated in this chapter, the consequences of this are as follows:

i) the transfer of cost from software tool to human resources (the low cost of software resulted in time-consuming, complex and to a large extent manual database maintenance)

ii) the transfer of workload from Consortium management to UDC MRF database administration (no management effort was put in data quality, product quality, documentation and procedures, or data preservation guidelines and procedures - all responsibility in these areas was left to the database administrator)

iii) the transfer of workload from the revision committee to database administration (revision output provides only a part of the data necessary for a database update, the notation validation and tracking is left to the database administrator, no editorial control over UDC MRF file distribution is established and the revision committee does not use or contribute to the full content of the database; for example it does not provide any input for administrative and data management fields)

iv) the transfer of cost of UDC use from the Consortium to its customers (a lack of data for implementation online and for the creation of tools, a lack of data for automatic update and a lack of more suitable exports cause costs that are entirely borne by purchasers of the UDC MRF)

When communication to users via *E&C*, the Web and IFLA updates are checked for the period 1993-2004, there was no record found of any business policy set up to alleviate the funding problem through an open invitation to users or the research community to exchange free input into the UDC MRF database in return for a UDC MRF licence or some other product. This research shows that the potential of the existing database is to a large extent unexploited. The existing UDC MRF could be easily used for creating better products or for providing better support to users at a very low cost or no cost at all. All existing UDC MRF annual files, for instance, could be easily put to better use in supporting users and improving or marketing the UDC.

Today, UDC is still one of the better-automated general classification systems with a range of quality classification tools built upon it. The reason for this is the initial work on schedule automation back in 1993 and the willingness of individual UDC MRF buyers to keep up with the additional costs of investing in this basic UDC data in order to create a more user-friendly classification tool. The number of actual system users drives these developments. The improvements to UDC data are, however, necessary and feasible for its use in a networked environment, and this will be addressed in Chapter Eight.

CHAPTER EIGHT: UDC DATA AND DATA MODEL

8.1 Introduction

This chapter looks at the relationship between database formats and the quality of UDC data. The data validation procedure and associated problems were examined using the 2003 UDC MRF.[126] This file was validated, both manually in CDS/ISIS and automatically in an external relational database, and the results were presented and analysed. This research is primarily concerned with a necessary improvement of the database format that would enable automated management and control of UDC data.

Describing the use of the UDC on the Internet in Chapter Three and subsequent chapters, it was frequently pointed out that there is a need for enriched UDC data: the need for hierarchy coding, an identifier, decomposed UDC headings, a subject-alphabetical index and mappings to other systems. In 1999 the Consortium considered the existing UDC MRF database, in order that it could support publication of the UDC MRF in an unlimited number of languages and, hence, be used for various multilingual editions. This created an opportunity for the improvement of the general UDC data format in many other ways, relevant for its use in information retrieval and its exchange in the networked environment. A suggestion for a new database format was made by G. Riesthuis in 2003, and is discussed in this chapter.

In order for the UDC to be used and shared in a networked environment, its data structure has to be transparent and clear, irrespective of the native database or target system of use. The required number of data elements and their function have to be identified from the point of view of their place in the classification structure, the function they have in indexing or information retrieval or the functions they have in UDC management and use. This chapter will, therefore, look at a conceptual data model as a common framework for understanding and implementing UDC as both a classification tool and in information retrieval.

[126] The UDC MRF 2003 contains changes introduced to the UDC at the end of 2003. The file was distributed in March 2004 and was valid until March 2005.

8.2 UDC MRF data validation

The period 1993-2004 included twelve annual updates and was a time of intensive change for the UDC system, and of innovation and experimentation with its structure. These UDC revisions introduced many new concepts and subjects, corrected mistakes in the hierarchy and reduced the enumeration of compound concepts represented by simple notation.

New classes were introduced and improvement was especially significant in the area of common auxiliaries (revision of Time, introduction of concepts of -02 Properties, and -04 Processes and relations). Some of the UDC changes affected the processing of UDC data, their quality and their amenability to automated control in the current database, and this section provides some evidence of the link between the database format and data control.

From the outset it should be clear that this research on validation makes a distinction between UDC MRF **data validation**, which can be performed and controlled both manually and automatically, and UDC **content validation,** which can only be performed manually and is outside the domain of database management.

Content validation is the responsibility of the UDC revision staff and covers several problems: *(a)* inconsistencies in style, formatting and the use of terminology (see Appendix 10, Table 10.1); *(b)* a lack of text structure in areas where patterns can be established; *(c)* confusing examples of combinations: incorrect syntax, valid but incorrect choices of notations to denote a concept[127] and incorrect descriptions accompanying examples of combination; *(d)* typographical mistakes that ought to be identified and corrected.

A need for this kind of content checking is illustrated below with an example of a record that contains two kinds of notes in the *scope note* field. The content shown is not a scope note and would be better accommodated if placed in the *editorial note* field and the *see also reference* field respectively:

---

[127] For instance, in the example of combination below, *612.1 Blood* from physiology is used in the combined expression 'attitude to blood transfusion' - instead of using the existing number *615.38 Blood transfusion* to express the subject more accurately*:*
2-446Medical care and treatment
Example of combination:
2-446:**612.1** Attitudes to blood **transfusion**

UDC-#[001]: 630 Table[002]: M 048970 DESCR[100]: Forestry SCOPE[110]: The subdivisions of 630 are a selection from the Forest Decimal Classification (IUFRO, Vienna). For fruit-tree growing see 634.1/.6

Content faults and inconsistencies checking, further illustrated in Appendix 10, Table 10.1, are outside the scope of database management and cannot be discovered automatically. The UDC research database tool, however, is indispensable in supporting this kind of revision work, as it enables field content searching, collocation of data, cross checking data sets and formatting/sorting of printouts for further work.

Data validation, on the other hand, is concerned with checking of the elements of UDC systems that have to comply with a set of rules (presence/absence of data, data type, structure and coding). Data validation is closely dependent on the database format and its ability to separate elements of data that have to be checked and logically related. The expression 'invalid notation', for instance, means that the notational element found somewhere in the UDC MRF file cannot be related to any UDC MRF record, either because it once existed and the record was cancelled, or because the notation contains mistakes (e.g. missing symbols, missing parts of numbers or wrong sequences of digits).

Because of the existing UDC Consortium management policy, tools and procedures, as explained in Chapter Seven, UDC MRF data are more susceptible to the following type of faults:

- invalid notation (in examples of combinations, parallel division instructions, application notes, scope notes and references)
- use of freely composed range notation and notation with three dots, in parallel division instructions and application notes
- empty caption fields[128]
- lack of data, such as a lack of parallel division instructions, a lack of references that could be inferred from the notes, a lack of index terms for

[128] This problem was reported by the author of this research in 2004 and was corrected in the next version of the UDC MRF (i.e. 2004 released in 2005)

searching elements of pre-synthesised notations

- misplaced scope or application note text
- mistakes in encoding (missing tags, 'swapped' tags, the coding of different kinds of content with the same tag, coding of notation and text in the same field)

These mistakes originate from three different sources: *(a)* mistakes introduced during the original creation of the database, *(b)* mistakes introduced in the process of revision by issuing incorrect or ambiguous update instructions and *(c)* mistakes occurring during the process of a database update. The validation of the 2003 UDC MRF, reported in this chapter, encompassed all notations that appeared in the MRF, including fields with notes in which the notation usually appears as part of an instruction.

8.2.1 Validation of the UDC in the CDS/ISIS database

In Chapter Seven it was made clear that the present CDS/ISIS database allows only manual checking of data and permits incorrect tags, empty fields or incorrect notation to be entered and saved in the database. Partial manual validation of UDC data is normally performed as part of routine checking of the UDC MRF prior to distribution. The list of common free text searches against the database and mistakes that can be discovered through this procedure is shown in Appendix 6.1, Table 6.1.

The additional manual checking shown in Table 8.1, below, illustrates the type of control that is feasible and can be performed in the existing database through free text searching, which simply examines the presence/absence of data in a field and involves quick visual control of content. It is common for most databases to contain mistakes, which is why database administration has appropriate mechanisms for their control and removal, usually in the form of guidelines, procedures or validation tools. As explained in Chapter Seven this is not the case with the management of the UDC MRF.

The type of mistake, illustrated in the Table 8.1 below, will not significantly impede the use of UDC in an online environment. Mistakes are listed here mainly to stress the importance of data quality checking and to illustrate the type of validation available through the CDS/ISIS database.

Table 8.1: Patterns of mistakes found through manual checking of UDC MRF

	RECORDS	CONTENT	TYPE OF MISTAKES FOUND
003 SPECIAL AUX. TYPE	17,493	CODED VALUE	3 records contain code 'D' by mistake
011 INSTRUCTION FOR PARALLEL DERIVATION	925	NOTATION	5 records contain notation followed by text that should be placed in subfield ^t 2 records have tagging mistakes only 109 records contain data in the target notation subfield
100 DESCRIPTION	66,000	TEXT	33 records are missing descriptions
105 VERBAL EXAMPLES	8,229	TEXT	3 records contain German and not English text
110 SCOPE NOTE	775	TEXT	1 record is missing tag ^e in front of the text 50 records (at least) contain application notes mixed with scope notes 10 records (at least) contain *see also* references 2 records (at least) contain editorial notes 2 records exist with complete text in brackets
111 APPLICATION NOTE	1,316	TEXT, NOTATION	3 records contain repeated parallel notation instructions in two fields 111 and 011 4 records contain instructions for parallel division which are not recorded in field 011 6 records contain scope notes mixed with application notes
115 COMBINATION EXAMPLES	2,341	TEXT, NOTATION	1 record has a missing tag 1 record (at least) has a number combination extended with three dots 2 records (at least) use symbols + and [ ] incorrectly
120 PARALLEL DIVISION EXAMPLES	9	NOTATION	9 records are missing tags that separate text from notation
125 REFERENCES	6,892	NOTATION	124 records exist with some kind of description or text[129]

Mistakes that are more significant for using UDC in an online environment are those related to incorrect notation. When creating online classification tools it is important to be able to relate two UDC notations, that is, to establish a link between two UDC records in a database. Therefore, the main question that remains is whether or not UDC notations that exist in the fields of notes, instructions, examples and references are valid.

---

[129] This can be considered to be a 'mistake' in relation to the other 67,000 records that do not contain any text associated with references and are supposed to inherit the caption from their respective records.

8.2.2 Sources of incorrect notation in UDC MRF

The main source of incorrect notation is update instructions which contain notations in references and examples of combination that are not checked via a database search prior to being published in the *Extensions & Corrections to the UDC* (E&C). In addition, every notation cancellation triggers a change in records where this notation is repeated (examples of combination, notes, and references). In the current UDC MRF these occurrences can be detected only through repetitive free text searching for a sequence of characters that mark the beginning of each notational element, which is time-consuming and is not considered feasible for the process of every annual update.[130]

The following two approaches in revision policy increase the risk of mistakes:

- an increased number of pre-synthesised notations in the UDC schedules
- cancelling a notation and reintroducing it again with different meaning (sometimes even in the very same update)

Two important changes in revision policy in 1993-2004 introduced a greater number of notations than could easily escape control in the present database. The first change is the removal from the UDC system of simple notations representing compound concepts. When a compound concept is cancelled, the way of expressing the same concept through the synthesis of simple notations is demonstrated in an example of combination.[131] One consequence of this is a greater number of examples of combination and thus an increased chance of these containing mistakes. The UDC MRF field of *combination examples* in the current database format, however, is not designed to control notation that consists of more than two elements and mistakes

---

[130] Apart from being time-consuming, free text searching of elements of pre-synthesised notation requires a good knowledge of rules for UDC synthesis since the procedure for this kind of checking is not recorded in the database documentation. In addition, if the notation re-used is altered (e.g. in parallel division) or is provided as part of a complex combination (combination with special auxiliaries) the simple notation checking demands a high level of expert knowledge of UDC.

[131] For instance, *789.3* [music] *For triangles* is cancelled and this concept can only be expressed through the following colon combination of two main numbers (*780.8 Music for individual instruments* and *780.634 Percussion instruments. Including: Triangles).*

cannot be easily discovered or automatically controlled.

The second change is an increase in the use of pre-synthesised notations in the main schedules and a new approach of introducing pre-synthesised UDC headings as records in their own right in the MRF. This is the case with, for instance, classes *2 Religion* and *94 History,* in which concepts in the main discipline hierarchy (i.e. the entity facet) are expressed as a combination of a main number and a common auxiliary number.

The changes in *History* follow a revision policy that focuses on the removal of parallel divisions from the UDC. This policy, if implemented consistently, may eventually affect main class hierarchies in ethnic grouping, palaeontology, languages, linguistics, industry, as well as seven hundred other UDC classes that result from parallel derivation.[132]

When it comes to the UDC MRF data management, this change leads to a growth in the number of pre-synthesised notations in the main UDC schedules; that is in main number field **001**. When introduced in the UDC MRF, these notations can only be marked as simple main UDC numbers (with code 'M' in field **002** *Table*), as the current database does not have an option implemented to code pre-synthesised notation in the main UDC number field - **001**.

UDC-#[001]: 26"652"	Table[002]: M	015999
SpAux[003]: D		
DESCR[100]: Religion of the Biblical period. Ancient Judaism. Old Testament religion		

The example above shows a record of *Ancient Judaism* which is expressed through a combination of the notation for *26 Judaism* and the time common auxiliary *"652" Ancient time*.

UDC-#[001]: 94(689.7)	Table[002]: M	066680
DESCR[100]: History of Malawi		

The example above shows the result of the revision of *94 History,* which introduces the history of individual countries, with combinations of notation for *94*

[132] In the course of this research no published documentation was found that states the plan, or timeline for the removal of parallel divisions from the system.

History and notations for place/country from the common auxiliary of place (marked grey in the example).

The UDC has clear number-building rules that are applicable for the whole of the system and, in principle, the UDC MRF need not list ready-made pre-combinations in order to make the system usable. Therefore, the increase of pre-synthesised notations listed in the main table can be seen as an editorial decision that is based on some kind of assumption with respect to the target users/audience, specific kind of collection, size of collection, or types of UDC application. Although the selection of ready-built combinations may be relatively small, there is still significant enumeration of pre-synthesised notations in the main schedules that follow different notation patterns and are not easily controlled in the current UDC MRF database.[133]

8.2.3 Problems in tracking notational elements

The main problem of controlling notational validity stems from the fact that many simple notational elements are re-used in pre-synthesised notations and the current database format does not provide a proper way of coding notational elements so that they can be automatically traced back to their respective database record and automatically validated.

UDC update instructions prepared by the Revision Committee provide only part of the information that is necessary for the UDC MRF maintenance. Notations that appear in the *application notes*, *scope notes* or *combination examples* are not accompanied by source notation used for building pre-synthesised notations. As a result, the database administrator cannot search for the supplied notation and validate it quickly and accurately.

---

[133] In the period 1993-2004, on several occasions (*94 History* and *2 Religion*) the combination of special auxiliaries with main numbers were introduced in the main schedules (outside special auxiliary tables) e.g. *94(411).01* or *233-158G* or *271.2-282.7-247.* This is explained as an exception by the editor in chief and is not anticipated for the future. In 2004, combination with special auxiliaries in *94 History* were cancelled and in 2005 *2 Religion* will also be revised to organize special auxiliaries in the special auxiliary tables and provide pre-synthesised numbers in the examples of combination. So the general rule with respect to pre-synthesised notation that is going to be followed in the future is to provide pre-synthesised notation in the field of *combination examples* and to reduce their number in the main schedules.

For instance, when a number such as *678.6.067.029.4* appears in *combination examples* or *application notes* it has to be traced back to *678.6* and *678.067* and *678.029.4* respectively, since the cancellation of any of these three notations will affect the example in question. In some classes where special auxiliaries may appear on different hierarchical levels it is very time-consuming for the database administrator to trace the appropriate special auxiliary table.

The situation may be even more complicated where a notation provided in *combination examples* is the result of a parallel division[134], for example *-035.676.332.066.3* which can only be valid if the following three notations are valid: *-035.67 Colorants. Dyes. Inks. Paints*, *667.633.2 Coatings according to kind of binder* and *66.066.3 Emulsion breaking*. Such duplication of work (i.e. tracing back to previously built numbers) already performed by authors of the UDC schedules causes delays and makes validation, which can only be performed manually in the CDS/ISIS database, unmanageable.[135]

An additional problem that impedes validation and data maintenance is the fact that authors of UDC schedules may point to notation (usually in notes and instructions) which is only 'approximate' and does not exist in that exact form in any UDC MRF record. For instance, it is not unusual to find notation extended with three

---

[134] Parallel division is the structural rule that allows for a previously developed hierarchy to be re-used in building new concepts without the need to re-enumerate the entire hierarchy. For instance, the table of common auxiliaries of languages (around 1300 languages) can be used to build numbers for ethnic grouping, or linguistics of individual languages or literature e.g. *=**111** English language*, *(=**111**) English people*, or linguistics *802.**111** English language (as a subject of study)* or literature 821.**111** *English literature***.** Created numbers consist of a base number and additional number(s) but in the process of building - the facet indicator is dropped and the resulting pre-combined UDC number appears as if it were a simple number. To automate the notation building fully, it is necessary that every number derived from a parallel division is marked as such and that the source notation, base number/facet indicator and target notation are declared in a machine-processable way.

[135] According to the explanation provided by the editor in chief, the UDC maintenance and update is organized assuming that the database administrator need not do any notation checking since this is anticipated as being done by the revision staff. Editor in chief explained that supplying of correct notation is a part of the reviser's job and no additional editorial control is necessary. Hence, no procedure is established for the actual verification of the final revision output checking through a database search: which data is checked, through which kind of free text searching, when or by whom. For this kind of vocabulary management it is very convenient that CDS/ISIS does not actually have any automatic validation built in and it allows for incorrect data to be saved and for updated to be completed irrespective the actual quality of the data.

dots e.g. *"38..."* or *3...* while the exact notation in UDC MRF is "38" and 3.[136] Similarly, the use of range notation that cannot be linked to any particular record in the database as shown in the following example:

```
UDC-#[001]: (=11/=8)            Table[002]: f        010303
--->[011] =11/=8. (Table 1c)
DESCR[100]: Linguistic-cultural groups, ethnic groups, peoples
EDIT [955]: File before (=11/=2)
```

In the example above one can see a record display of a UDC number *(=11/=8)* containing a parallel division instruction in field **011** (marked in grey). This particular record exemplifies three different faults with respect to notation presentation:

- The instruction means that the UDC notation *(=11/=8)* in the respective record, shown in field **001** can be further hierarchically subdivided in the same way as the notation listed in field **011**: *=11/=8*. The notation that appears in field **011**, however, does not exist as a record in the UDC MRF - the correct UDC heading that exists in the UDC MRF would be *=1/=8 Natural languages*. The instruction cannot be automatically linked to the beginning of the hierarchy that has to be displayed in the process of number building
- The field should contain source notation and target notation, and only the source notation is present in this record
- In principle every parallel division instruction should contain a parallel division example. This is not the case with this record
- Field **011** provides separate subfields for notation and text and in this case text '*Table 1c*' is entered as a part of the notation.

Because CDS/ISIS functionality allows for free text searching against each field and because in the majority of cases, UDC notation contains symbols that indicate that a number is a pre-synthesised notation ,it was assumed that this should be sufficient for finding all the occurrences of a given notation. The possibility of tracking notation in the UDC MRF, however, is dependent on the manual searching for notation against

[136] In CDS/ISIS this notation can only be searched if enclosed in pointed brackets < >.

every field in which notation may occur (see Footnote [130]). This kind of manual validation assumes that the database administrator will know, for instance, which notation in examples of combination or elsewhere require splitting and into what notational elements.[137]

Taking into account the fact that the current CDS/ISIS database does not provide any automatic field validation and the database format does not allow for coding of notational elements, it is realistic to expect that in the process of 12 updates (1993 to 2004), a certain number of incorrect notations were entered into the database and a certain number of cancelled notations still exist in the examples of combinations, references and note fields.

8.2.4 Validation of notation using a relational database

Any simple UDC notation appearing in fields other than **001** was considered valid if it could be matched to notation held in field **001**, that is, if this notation has its own record in the UDC MRF. Likewise, simple notations from other fields in the database that could not be matched to notation contained in field **001** were considered invalid.

For instance, if a notation such as *-052.3,* which is a simple notation from the *common auxiliaries of persons* table, was found alone or in combination with other notations anywhere in the UDC MRF (scope notes, application notes, references, examples of combination etc.) this notation would be considered valid as it can be matched to notation contained in field **001** of the following record:

```
UDC-#[001]: -053.2                Table[002]: k        011904
DESCR[100]: Children and infants (in general)
REF->[125]: -055.15; -055.25; -055.62
```

Similarly, if a special auxiliary such as *-006.3* is found anywhere in the UDC MRF combined with UDC notation starting with *61*, such as *617-006.3* or *611.77-*

[137]Riesthuis, for instance, finds that his program for the decomposition of UDC numbers can be used for splitting the majority of pre-synthesised notations in the examples of combinations. Once this is done numbers can be manually searched. This program, however, is not a part of the Consortium database management tool and is not available for use in a UDC MRF update.

006.3, it will be valid as it can be matched to a record in the *special auxiliary table*:

```
UDC-#[001]: 616-006.3          Table[002]: M      037063
 SpAux[003]: A
 DESCR[100]: Tumours derived from non-haemopoieti (non-haematopoietic) mesenchymal tissue
```

The notion of validity is linked to the possibility of tracing the re-use of any simple UDC notation throughout the database. In terms of database management this tracing means that if any simple UDC number is cancelled, and its record is removed from the database, all the occurrences of that number throughout the database can be found and removed. If this is not the case, the database will contain invalid numbers.

The complete validation of the notation is performed by importing the UDC MRF into a relational database called the *UDC Research Database*.[138] After relational tables relevant for UDC were created, it was possible to search and find all notations that could not be matched to existing UDC MRF records. The important part of this research was to find out why invalid numbers occur, how easy it was to find all simple notational elements re-used in different combinations and what changes in database format would be needed to prevent mistakes from occurring.

8.2.4.1 Notation validation findings

The list of fields included in the validation of the UDC MRF 2003 is given in Table 8.2. All the fields included contain pre-synthesised notations held in the same format as simple notations. For the purpose of validation, pre-synthesised notations had to be extracted and split into simple components. This can be accomplished automatically only for notation that contains symbols indicating the beginning or ending of the notation. In the case of notation containing combinations of special auxiliaries and parallel divisions this had to be done manually, as the present database does not contain the data necessary for automating the process.

---

[138] The MySQL database used for this purpose was originally designed to support a faceted classification tool (SOFathum) created in 2003 for the FATKS project by the author of this research and a software developer (http://www.ucl.ac.uk/fatks/fathum.htm). For the purpose of the UDC validation, the author of this research has restructured the database according to the UDC conceptual model and UDC entity relationship model, described later in this chapter. A software developer helped by creating additional C++ programs for data import and more complex validation.

Table 8.2: Fields included in automatic validation

FIELDS VALIDATED	FIELD CONTENT	OCCURRENCE
001 UDC number	notation	NON REPEATABLE
010 Parallel derivation (derived from)	notation	NON REPEATABLE
011 Instruction for parallel division **^a** Source notation for the parallel division **^b** Target notation (if not identical with the number in field 001)	notation	REPEATABLE
110 Scope note	text mixed with notation	REPEATABLE
111 Application note	text mixed with notation	REPEATABLE
115 Combination examples **^a**Direct addition **^b**Colon (:) combination **^c**Full notation	notation	REPEATABLE
120 Parallel division examples	notation	REPEATABLE
125 References	notation	REPEATABLE

Range numbers were split automatically into two numbers: the first containing the beginning and the second containing the end number of the range, for example *539.122/.126* was validated as *539.122* and *539.126*. This was assumed sufficient for the purpose of this specific research, whose only purpose was to illustrate that mistakes can occur even in the beginnings and endings of ranges which are clearly presented and recorded and can be searched for in CDS/ISIS.

Special auxiliary numbers in the *combination examples* field represented an obstacle to validation, as they could not be traced back automatically to their table. For instance, example *784.087.68.*092 *Competitions for choirs* contains main number *784 Vocal music* combined with two special auxiliary numbers, one from the special auxiliary table at the beginning of class *7*, i.e. *7.092 Competitions* and the second from the special auxiliary table in class *78*, that is *78.087.68 Choirs*.

The UDC MRF database does not hold data on the 'source notation' associated with *examples of combination* and tracing had to be done manually. It was not feasible to do this for all special auxiliaries appearing in the *combination examples* fields and this part of the validation was not complete.

Finding and extracting notation from the *application notes* and *scope notes* fields was also difficult. Notations found in these fields were mixed with text and consisted of simple and pre-synthesised notations but also contained extensions with three dots. The procedure was first to extract the notation from these fields, then to separate simple from synthesised notations and split the synthesised notations into components that need to be validated.

The result of this validation is summarised in Table 8.3 and a full report on incorrect notation for each field is provided in Appendix 10.1. During the process of a regular UDC MRF update, the main UDC number field **001** is the most rigorously checked field since this field is used for sorting the database. No invalid notation was found in this field.

It is important to point out, as shown in Table 8.3, that field **001** contains almost 18,000 pre-synthesised numbers, and over a thousand of those are combinations with common auxiliaries. Of the total number of pre-synthesised notations, there are 736 range numbers, 6 colon combinations, 20 notations containing + (all used in common auxiliaries of place), 228 combinations with common auxiliaries of place, form and ethnic grouping and 21 combinations with time auxiliaries. There are 132 pre-combined numbers combinations with -02, -03 and -05 that, although they have the form of common auxiliaries, are actually special auxiliaries from class *61 Medicine.*

The field of *parallel derivation* **- 010** contained only one invalid notation out of a total of 902. The field of *parallel division instruction* **- 011** contained 925 notational elements with 19 invalid numbers and 100 range numbers that could not be matched to any record in the MRF.

In the *scope note* field - **110** there were 165 pre-synthesised notations (68 were combinations with special auxiliaries, and 97 were others). In spite of the nature of this field, which is supposed to contain only text, there were 708 notational elements with 28 incorrect notations. In *application notes*, logically there were many more notational combinations (622). When split, these contained 1,482 notational elements containing 21 invalid notations when matched against the UDC MRF.

Table 8.3: Automatic notation validation

FIELD	CONTENT	NUMBER OF RECORDS	PRE-SYNTHESISED NOTATIONS	NUMBER OF NOTATIONAL EL. VALIDATED	CANNOT BE LINKED TO UDC MRF RECORDS
001 UDC number	notation	66,733	1011 (+ 17,473 in special auxiliary tables	2, 602	**0**
010 Parallel derivation	notation	902		902	**1**
011 Instruction for parallel division	notation	925		925	**19** (+ 100 range numbers cannot be matched with UDC record
110 Scope note	text/notation	775	165	708[139]	**28**
111 Application note	text/notation	1,316 records containing 1,377 notations)[140]	622	1,482	**21**
115 Combination examples	notation	2,341			
^a Direct addition		1,023	437	1,184	**55**
^b Colon (:) combination		935	366	2,164	**58**
^c Full notation		384	797	1090	**48**
120 Parallel division examples[141]	notation	9	----	15	---
125 References	notation	6,892	104	10,323	**394** (30 wrong notations and 366 number combinations that cannot be linked)
TOTAL: 1203 invalid notations					

The field of *combination examples* (**115**) contains three subfields. The direct addition subfield ^a contained 1,184 notations, 437 of which were already pre-combined. Validation revealed 55 incorrect notations. Subfield ^b colon combination, contained 2,164 notational elements (366 were already pre-

[139] 32 notations in this field contained 'three dots' which had to be removed in order to validate the number, on the other hand 79 notations contained a range number and these had to be split to be validated.

[140] Pre-combined notations that had to be split prior to validation were combinations with common auxiliaries, 602 notations contained 'three dots' which had to be removed prior to validation, 39 notations contained a colon (these were split into simple notations) and 13 notations contained a plus symbol, which was removed. Notations that contained range numbers were split into two notational elements.

[141] Notation appearing in this field need not be automatically linked to the records of the originating notation (this is the purpose of field 011).

synthesised[142]), 58 being incorrect. The last subfield ^c contained completely built notations. When split, these amounted to 1,090 notational elements with 48 incorrect notations. Examples of combination with 4,438 notations and 161 invalid notations proved to be the most time-consuming to check as this field contained a great number of pre-combined numbers that had to be split manually in order for them to be matched automatically (see report in Appendix 10.1, Table 10.5)

In addition, because of the fact that the majority of special auxiliaries were not validated, the actual total number of incorrect notations is likely to be somewhat higher. The *references* field contained over 10,000 notations. This field's primary function is to connect two MRF records and with this in mind the validation was purposefully completed automatically and therefore quick to perform. It showed that 394 notations cannot be connected to or do not point to any MRF record. When analysed, references contained 30 incorrect notations (see report in Appendix 10.1, Table 10.6) and 366 pre-combined numbers, i.e. examples of combination rather than simple UDC numbers, which is why they could not be linked to any UDC MRF record. In addition, automatic validation of this field in the relational database established that half of the references (5,591) were not cross-referenced at all.[143]

The extraction of notation from the fields listed shows that pre-synthesised notations occur in all fields and that this kind of notation is not easy to find or control automatically. When pre-synthesised notations were broken down into constituent elements, it was revealed that there were at least 21,397 simple notational elements (or special auxiliary combinations) repeated in different fields of the UDC MRF which have to be traced back to their original database record. The number of pre-synthesised notations indicates a problem in the management of data caused by the limitation in database format that does not provide a sufficient amount of coding elements necessary to manage pre-synthesised notation or to trace special auxiliary tables and parallel division instructions.

The existing database relies on manual editing of each notational element in

---

[142] This subfield contained 182 combinations of main numbers with special auxiliaries and 184 combinations of main numbers with common auxiliaries or other main numbers.

[143] For instance *(01) Bibliographies* points to *016 Special bibliographies* but 016 does not point back to *(01) Bibliography*; or *(03) Reference works* points to *030 General reference works* but 030 does not point back to (03).

field **951** *Index only UDC notation* at the end of the editing worksheet. This approach does not seem to be working as only 566 records in the UDC MRF have this field populated and only 32 were added during the period 1994-2004, that is, after the creation of the database. The total number of indexed notations is 1,084 numbers. A more intuitive approach would be to code notational elements immediately next to the field where pre-synthesised notation appears.

The significance of the total of 1,203 notational elements (Table 8.3) that could not be matched or linked to the UDC MRF is related to the way the UDC MRF is used. If a database file is sold for the purpose of translation or a printed edition, the problem will be transferred to publishers who may discover and remove the incorrect notation during the process of translation or preparation of the edition. If a database file is purchased for direct online use, publishing electronic editions or building classification tools, the aforementioned mistakes in the notation may cause delays and a change of the original plan with regard to tool functionality (i.e. giving up the hyperlinking of notes, examples or references). Once reported, however, mistakes do not create a problem and can be corrected if the Consortium or revision committee wishes to do so. For this research it is more important to point out the difficulty in finding, controlling and preventing mistakes from appearing in the present database.

The validation exercise undertaken in this research demonstrates that if one wishes to improve UDC MRF data in the current database, this would still have to be done manually and there is no easy way of preventing mistakes appearing in the future without a change of database format. The problems for automatic processing and validation may be caused by variations in presentation of textual data in *scope notes* and *application notes*, which contain embedded rules, instructions, examples of combination and even references. The UDC MRF was created from a printed edition. According to the way authors of the schedules provided textual notes and treated notation in the field of references, as well as the way schedules were presented (cf. Appendix 10, Table 10.1) in the period 1993-2005, it is obvious that the UDC MRF continues to be revised with a printed edition in mind. Adherence to this policy indicates that there is a need for the database format to be changed in order to separate further the structured data necessary to manage and exchange UDC online, and the unstructured textual output of the UDC revision.

For fields that contain only notation, the most obvious problem is a lack of data that would enable the coding of pre-composed notation. Simple notation in a field that can be established as notation-only, such as **125** *References*, can be easily validated with an external tool and it is only pre-synthesised notation that is not properly encoded that poses a problem. This validation shows that all fields in the UDC MRF that contain notations would need to have a 'notation validation subfield' which could be used to declare the base notation in cases where the UDC string is a combination of special auxiliaries, is the result of parallel derivation or when notation is altered in style and presentation. The advantage of having these data as part of the field in which the pre-composed notation is stated, is the ease of visual control in the process of database editing.[144] Only by gradually enriching the UDC MRF with this kind of information would it be possible to establish full control and track variants of UDC numbers in different combinations.

The weakest points of the present database format are the coding of special auxiliaries, examples of combination and pre-synthesised UDC headings. The problem is simply the lack of data necessary to control and manage this part of the UDC properly. This can be only changed by improvement of the database format. In addition, this validation research has confirmed the need for changes in UDC notation, some of which were reported earlier by Riesthuis (2001, 2003):

- a differentiation between the apostrophe used for the shortening of notation and the apostrophe used as a facet indicator in special auxiliaries[145]
- a change of facet indicator for the 78 special auxiliaries in Medicine starting with -03 as its coincides with common auxiliaries of materials (this will be removed with the currently ongoing revision of Medicine)
- revision of 200 main numbers containing .0 notation as this coincides with special auxiliary tables
- considering the formal change of facet indicator for the common

[144] In addition, it is paramount to improve the revision procedure and output in order to record the original notation in the process of number building, in the case of examples of combinations containing special auxiliary numbers and numbers derived from parallel division. This would prevent unnecessary duplication of work and would speed up the update procedure.

[145] In the UDC MRF (2002), in order to differentiate between the two, an inverted single comma ` was introduced for special auxiliaries. Riesthuis' opinion is that the apostrophe used for shortening of notation should be completely abolished.

auxiliary of time from straight-up typewriter quotes "..." to symbols with different beginnings/endings such as inverted commas “...”.[146]

8.3 Riesthuis' proposal for the new database format 2003

Since its creation in 1993, the MRF database has not been developed further in terms of structure or accompanying tools. The database was designed with the UDC data structure in mind that was valid in 1990-1993. It was anticipated that the database tool would be used in the revision process and that this would lead to an incremental improvement of both the database and UDC data. This was not the case and no work on systematic checking, improving or enriching of the database structure or data coding took place in the period of 1993-2004.

In 1999 the Consortium considered the possibility of a multilingual UDC MRF at which point it became obvious that the present database format, created to hold only English and German text, had to be adapted for the purpose. In relation to this request, the first consideration towards the new database format was put forward by Riesthuis in 2001, revised in 2002, and the final proposal was made available in 2003 (Riesthuis, 2001, 2002, 2003).[147]

The previous section on UDC MRF data illustrates existing problems of validating UDC pre-synthesised notation in the current CDS/ISIS database. The shortcomings of the existing database format and problems in managing UDC data and their use in an online environment were well known to the designers of the original format and to the Consortium and were reported in the literature (Riesthuis, 1994, 1998, 2001). The problems were caused by the fact that textual data used to produce UDC MRF did not allow for the automatic population of all data elements that are necessary for UDC management online. When proposing a new restructured database Riesthuis (2001) emphasised the following known problems that have not been solved in the original database design since 1993:

- a lack of data linking each main number to a special auxiliary table

[146] The change of "" to «» (ASCII 174/175) has already been done in the UDC MRF and it is very easy to change this to inverted commas. There is no reason not to make this change a more official alternative solution for "" on a system level.
[147] Riesthuis' suggestions coincided with the extreme difficulties in the Consortium finances and the proposal is still awaiting due consideration.

with which it can be used

- a lack of data on a correct hierarchy
- a lack of data on the type of UDC number (simple, synthesised, consisting of which elements)
- a weakness in language coding within textual fields (and the inconsistency in coding)
- a coding inadequacy in auxiliaries of time that should contain different beginning/ending symbols instead of "..."

(Riesthuis, 1998, 2001)

The full extent of changes planned for the new database format in the latest and revised proposal by Riesthuis (2003) is shown in Appendix 6.2 and this is compared to the existing database element schema. Because of the need to manage repeatable language subfields, most existing fields had to be restructured. In summary, the new database proposal contains the following important improvements:

- the introduction of language subfields for all textual fields (23 fields)
- the introduction of four new fields:
    - 004 Combination type[148]
    - 005 Hierarchical next higher notation
    - 006 Special auxiliary valid for 001
    - 126 Index terms
- the removal of nine fields
- a change of facet indicators for the main auxiliary of time

There are only six fields that were not affected by the new database proposals (**001** *UDC number*, **002** *Table*, **003** *Special Auxiliary Type*, **010** *Parallel derivation*, **012** *Special auxiliary supplied by parallel division*, **013** *Special auxiliary supplied by annotation).* Four important new elements were suggested for addition and the subfield structure was proposed to be changed in twenty-three elements.

All administrative elements were proposed to be restructured and nine

[148] This field was proposed in 1993 but has never been implemented.

elements to be removed and replaced with adequate subfields in the remaining restructured fields.[149]

Apart from additional fields, one of the most important improvements in terms of database management would be the implementation of some of the functionalities in data validation and editing that are available in WinISIS. Riesthuis proposed the implementation of 'pick-up' lists, but most importantly, he suggested mandatory fields and conditional mandatory[150] fields, which would help control data population. It is hard to introduce more rigorous and more extensive data validation supported by WinISIS because of the present state of the UDC MRF, which continues to be updated with roughly structured UDC data. As long as the revision output remains an MS Word document formatted for publication of the *E&C,* containing only a percentage of the total data necessary for full data coding, it is difficult to implement any further automatic data-quality control.[151]

For the creation of online tools using the UDC, the most urgent of all problems to be solved with a new database proposal, was with automatic linking of special auxiliary tables to their place of application. Riesthuis gives an example of class 6 where more than one special auxiliary table is applicable throughout the classes: "in the record *616* is a note saying that the special auxiliaries given under *616* are valid also for *617* and *618*. Under *617* and *618* there is no reference to these special auxiliaries" (Riesthuis, 1998: 24). As shown in Figure 8.1, the link exists in a top-down direction and is missing a bottom-up relationship that makes the automation of linking impossible.

To address this problem Riesthuis proposes the introduction of a new data element: **006** *Special auxiliaries, valid for the notation in field 001*, which would contain, for every UDC number, a pointer to a special auxiliary table that can be used

---

[149] Fields removed are: **903** *Introduction Source,* **904** *Introduction Comments,* **912** *Replacing UDC Notation,* **913** *Source of Cancellation,* **914** *Cancellation comments* were removed **921** *Last revision date*, **922** *Fields revised,* **923** *Source of revision*, **924** *Revision comment*.

[150] This allows the linking of the presence of one field to the population of another.

[151] The current revision output has very loose structuring rules as it was created for publishing the E&C, and has no relationship to the structure of the UDC data. Often, this text is inconsistent with the data field semantics. For instance, the output often confuses application, scope or editorial notes or will simply mix all three. References will usually appear as simple notation but occasionally they will contain some textual note. Examples of combinations appear in UDC filing order while in the UDC database they have to be held according to the type of combination.

for its further specification. If implemented this would be an important improvement of the UDC data and would enable automatic linking of each UDC number with its special auxiliary table(s).

One issue that is not going to be solved simply with the proposed field relates to a lack of data in the present MRF indicating which of the 17,493 records containing special auxiliary numbers contain *the rule*, i.e. which of these records marks the beginning of the special auxiliary table.

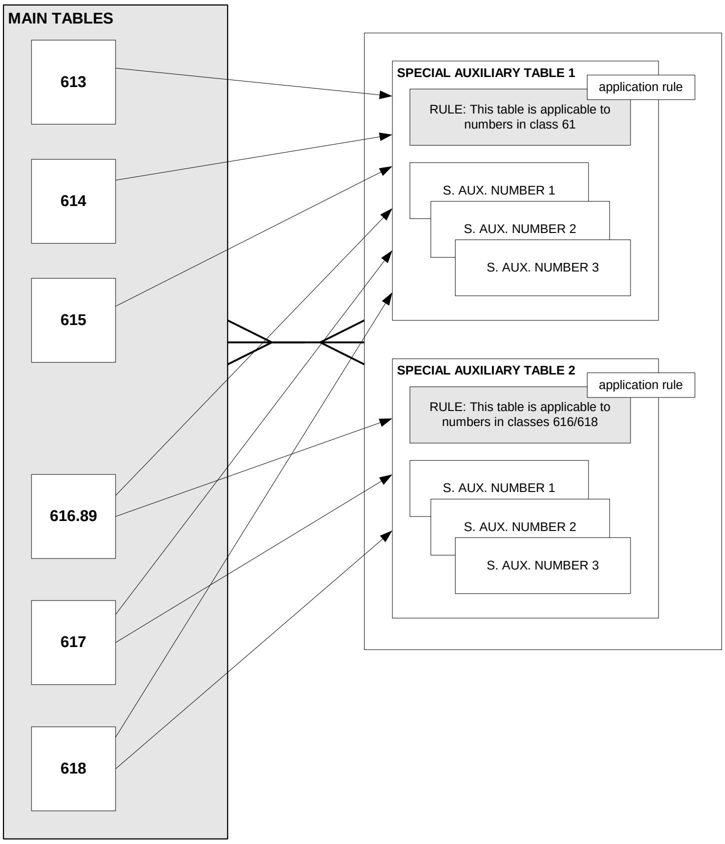


Figure 8.1: Relationships between special auxiliary table(s) and main table

At the moment there is no possibility in the UDC MRF to establish automatically how many auxiliary tables there are in a certain class and how many there are at a system level. The introduction of a new **006** field would not address this problem. This could be solved by adding some kind of simple code in the existing field **003** *Special auxiliary type* to flag every record in the special auxiliary table that denotes the beginning of the table and contains the rule for its use. At the moment a similar function is provided through field **013** *Special auxiliary supplied by annotation* which marks records containing application notes relevant for special auxiliary tables, but this information is not provided in the same way for all tables within the system. What makes the management of special auxiliaries more difficult to automate in online tools is the incoherence with which the tables are announced, introduced or accompanied by application notes (illustrated in Appendix 10.1, Table 10.1).

Identification of special auxiliary tables would help in the management of the revision process and in controlling the number of special auxiliary tables in the system. It could also be a useful control in preparing schedules for printing but most importantly, together with field **006**, it would serve in online UDC editions for automating the linking and display of tables next to each UDC number.

Riesthuis acknowledges the growth of pre-synthesised notations in the **001** UDC number field and the problems in the management of individual elements. He suggested, however, that field **001** *UDC number* should not be changed or structured because of the complexity of notation in cases such as *745/749'02/04'* or *745/749(492/493)* (Riesthuis, 2001: 17). The possibility of introducing an alternative field that would hold fully coded pre-synthesised UDC numbers was not taken into consideration in this database redesign. In addition, it is unfortunate that no change was proposed for field **115** *Examples of combination*, which was designed to manage properly the direct addition of a simple number and colon combination with a simple number but does not provide any solution for the situation when a pre-synthesised notation follows the colon or is directly added to a main number notation. It seems that there is no real benefit in having three different subfields when these do not enable a search of individual elements.

Combined with proper language coding, the implementation of the new database format (as proposed by Riesthuis) would bring substantial improvements to the UDC MRF. In relation to the requirements for information retrieval and UDC implementation and management, as discussed earlier in this thesis, and particularly in connection to data validation, it is, however, evident that several issues will remain unresolved in the proposed new database. Some of the data fundamental for UDC maintenance, exchange and use are still missing. The following improvements remain to be considered for the future.

i) introduction of a **unique UDC number identifier** to support quicker and more accurate automation of updates for the purpose of vocabulary exchange[152]

ii) introduction of a field for a **structured UDC heading** (in parallel to the existing one) to support proper management of the UDC in the UDC MRF database and later on in its IR implementation

iii) **restructuring of the combination examples field** to contain fully coded number building elements so they can be properly searched and maintained

iv) improvement of the **notation indexing** for fields that contain pre-synthesised numbers (scope notes, application notes, examples of combination, and parallel division examples) for the purpose of UDC MRF maintenance

v) introduction of a 'flag' code for the beginning of a special auxiliary table i.e. the record containing the table application rule - so that special auxiliary tables can be properly managed and controlled in UDC MRF maintenance and in the creation of UDC classification tools online

As was explained earlier in this chapter, incorrect notations are introduced into the UDC MRF through unchecked and non-validated notations in the process of revision updates, over which the database administrator has no control. This means that it would be better if data quality in the database did not rely solely on the goodwill of contributors but on the automatic validation and upgrade of the database to a tool that would not permitting incorrect notations to be saved. With the proposed database format and no changes in management of updates or changes in the format of revision output, however, it would still not be possible to stop invalid notation

[152] Ideally this would be a purely numerical code for quicker and easier machine processing. UDC numbers are not suitable for this purpose for the following reasons: *(a)* UDC numbers are occasionally cancelled and re-introduced to denote different concepts, *(b)* the UDC MRF file does not contain cancelled numbers and in conjunction, the database record number is changed with every update. In addition, UDC notation contains symbols and letters, which make it inefficient for machine processing. The scenario for exchange and access of the UDC in the networked environment using a URI in the form of: *info:kos/scheme/«code»/«expr»/«lang»* and *info:kos/concept/«code»/«id»*, as proposed by OCLC for their terminological services, may be further explored only if the policy towards the UDC unique identifier is fully implemented on the level of UDC data source.

from appearing in the database nor to detect these mistakes automatically prior to database file distribution.

During the course of this doctoral research the comments on the database proposal were sent to Riesthuis, who had to make a choice between various options in order to make the new database project feasible. The Consortium should now respond to the Riesthuis' suggestion and speed up the new database implementation.

8.4 Re-thinking the rôle of the database tool

The first UDC MRF database in 1993 was created primarily to enable maintenance and publishing functions. In the subsequent period from 1993-2004, the UDC has also been used, published and exchanged in different applications online and this kind of use has exposed more requirements with respect to the total number, formality and transparency of UDC structural data elements. The UDC MRF is the main source of UDC system data and it is anticipated that this database file does not contain less information than is necessary for the classification to be fully used in a given environment.

For this research, focusing on the potential and future use of the UDC in the networked environment, the strength of the UDC database format is paramount. A well structured and robust database format will ensure a quality source of UDC data that will guarantee a more sustainable development of the classification and its wider application scope.

The amount of data in the version of the MRF distributed to users as a text export, called UMRF, for instance, is not sufficient for online applications. Users usually need a text export that contains data on tables, editorial and other administrative notes. In order to have access to these data, users have to use the full UDC MRF format in a text or ISO export since no other more suitable format is made available for distribution. In order to use the UDC with a user-friendly interface and have the possibility for browsing the classification, users usually create a relational database. For this to work, they need to build middleware tools that will convert the UDC MRF into their own database format. This scenario is a reality for numerous users or publishers who have been working on translations, preparation of CD-ROM tools or publishing the UDC online in the period 1993-2005.

With this in mind it would be beneficial if the Consortium considered the following two points seriously. The first is that the proposed database format, which would provide enriched UDC data exports that are cheaper to implement, is well overdue and that steps should be taken for this to be realised as soon as possible. The second is the fact that the change of the UDC MRF database structure and its export formats will greatly affect licence holders maintaining online applications, and is likely to disturb updates and inflict unwanted costs and delays. It is, therefore, important that the database change is managed responsibly and professionally, with an official plan and documentation on database changes communicated to users well in advance (i.e. one licence period). This has to appear together with the database independent structured text export(s).

The new database 'neutral' export format is necessary in order to bridge the old and new database structure which means that users adopting this export format will not be affected by the database change. The structure and data elements of this format ought to be made public and accessible to current and potential users so they can re-design their tools.

Up to 2004, the Consortium was hesitant to consider the need for more diverse and improved export formats in line with real-life use of the system. It is, therefore, important that the Consortium markets the improved database format and that it expresses interest in securing the place of the UDC in the networked environment through improving interaction with users.

There is no real obstacle to the UDC MRF database becoming a more versatile tool and database management system, which would suit a greater number of different publishers' and users' needs better. When planning a database redesign, one of the shortcomings that can be avoided is a lack of mapping elements in the database schema. The present and proposed database formats do not provide the possibility for mapping to systems outside UDC or for possible development of thesauri or any other subject-alphabetical indexing system.

Similarly, there is no provision for handling different editions of the UDC that may be held in the same database format such as special, abridged or extended applications or specific collection applications. These data elements, if implemented, would make the UDC database format suitable for meeting both the present limited needs of the Consortium and potential needs of individual Consortium member projects or licence holders.

It is also possible for a classification database tool to be created not only for publishing and distribution but also in accordance with the prospect of more advanced UDC revision plans and tasks in mind. In such a scenario, it is possible to envisage that in future, revision may be more focused on the improvement of UDC data in terms of structural coherence and content quality as well as more rigorous data formality. This would be, for instance, in line with the current trend in developing ontology-like vocabulary tools that would be more formalised and hence more amenable to machine processing.

If this were to be the case, then one could consider the introduction of data elements that might be of less interest to current users or publishers but may be very useful in managing the classification structure. One example is information on the type of semantic relationships (type of hierarchical relationships generic/partitive, class/member, class/instance, or type of associative relationships).

In addition, if it is envisaged that the UDC system will be revised to incorporate a fully faceted structure, apart from the proper guidelines for the implementation of a faceted structure in line with UDC notational and syntagmatic rules, it would be useful to have a mechanism for controlling and coding fundamental facets (entity, kind, part, processes, operations etc.). The coding of facets is an important step in the preparation of work in facet analysis, planning and restructuring of the existing vocabulary.

All existing UDC applications, whether in preparation or in final delivery, are tied to the machine processing of UDC data. A lack of data in the UDC source format can be very limiting if not crippling when it comes to potential and future use of the system. Figure 8.2 shows the scope of what could be called a desirable 'UDC service model', which corresponds to the needs of the present UDC use environment.

Any implementation of the UDC in an online environment is concerned with a number of functions that depend on the availability of properly structured data. Many users need to do some data improvement in order to make their UDC application work online with the present UDC data format. Although the UDC implementation costs have always been transferred to users and the UDC Consortium continued with the same tradition in 1993-2004, this may need to change in the future if UDC is to remain competitive in the KOS tools market.

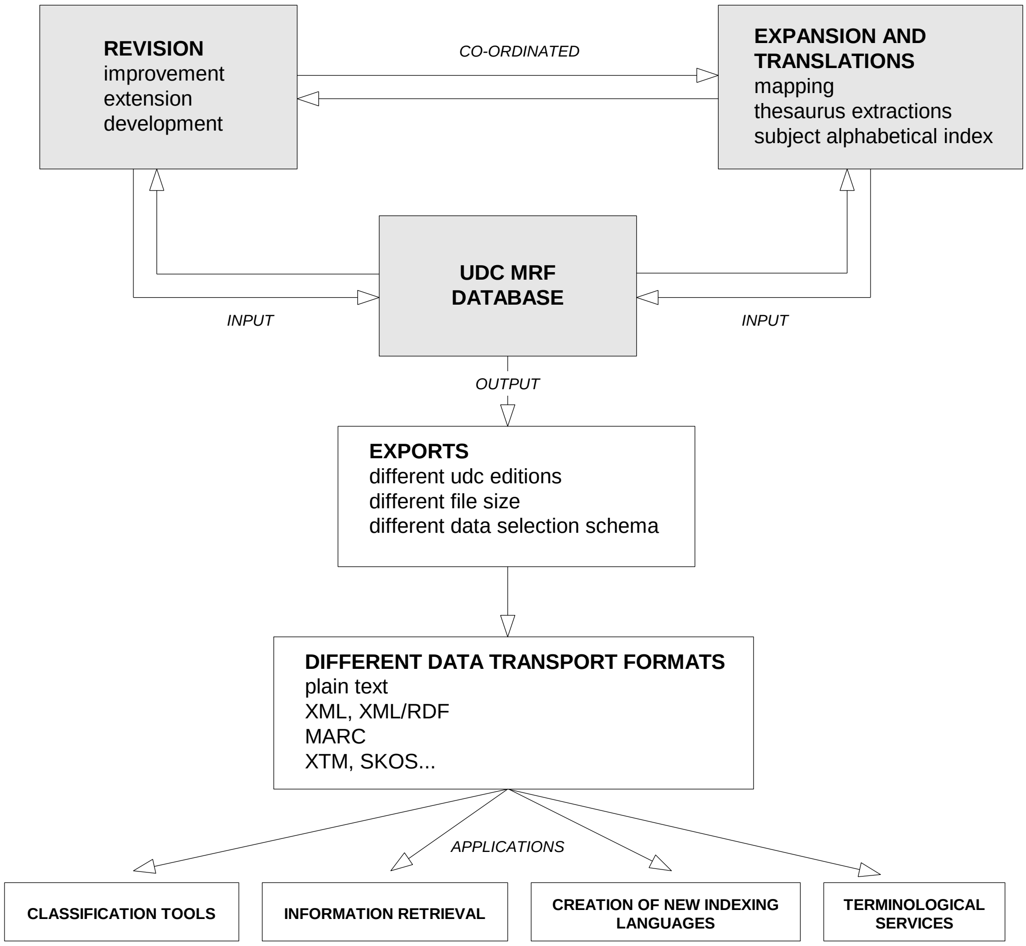


Figure 8.2: UDC database service model

Improving and enriching UDC source data, although completely outside the current owner's agenda, may have far reaching effects for the suitability of UDC in a networked environment. Database format improvement is the first step towards any kind of future system development.

8.5 Proposal of an improved UDC data model

In many cases the cost of UDC implementation is increased because of the time online tool implementers need to understand the underlying data structure and the way it functions. This was clearly the case with the creation of the BSI *UDC-Online* tool in 2001 and the re-building of the tool in 2004 (Appendix 6.2). While the cost of *UDC-Online* is recoverable through the product's commercial value, this is not the case with most UDC applications in information systems that are not commercially orientated, such as library systems, digital libraries or resource discovery on the Internet in general.

If an online UDC application aims to support searching and browsing UDC notation and/or captions and parallel facet browsing in a user-friendly way, the application would benefit from the functions provided by a relational database tool. The UDC MRF Manual (2003) is the only document available on the UDC data format and does not discuss the data model on which the UDC MRF format was created, which would be useful in creating online applications.

Also, the existing database format, created for CDS/ISIS, does not provide sufficient information on relationships between data elements or their functions. This information on UDC system structure would help the conversion of the UDC MRF into a relational database for implementers not knowledgeable about the UDC (see Appendix 10.4). The mixture of notation presentation style, unclear separation from textual and purely notational data elements and the number of pre-synthesised notations that cannot be related to existing database records complicates the implementation further.

The model of UDC data in IR systems, for instance, can be simplified by treating a combination of special auxiliaries and parallel derivations as simple concepts. Information loss can be prevented by providing: pre-built combinations, relevant verbal access points and semantic linking between source concepts and created compound concepts. There are no guidelines for users and developers of online applications that would help in the creation of simple and functional models and transparent representations that can be linked to the 'flat' UDC data currently distributed.

The choice of a classification system for potential users of the UDC is driven by the transparency of the data structure and the ease of its implementation. The more transparent the data model, the easier and cheaper it is to implement. The UDC MRF does not have sufficient data to make the UDC easy to implement, and its data format makes it more difficult to see through the full and yet unexploited potential of the classification structure and its content.

Information about UDC system rules and structural elements would be useful in a more universal database language, for anyone implementing the system online. A UDC conceptual data model, i.e. a conceptual schema, independent of any specific source or target tool, could play an important rôle in online implementation and could be used for different purposes.

A conceptual schema is supposed to guide the development of the future

UDC MRF database management system, the development of classification tools, data migration or data exports. It may be useful in helping improve, not only the use of UDC in existing IR systems, but also systems outside the UDC-specific domain where it could be used as a test-bed for creating standards and formats for classification exchange. Such a conceptual schema is proposed here and shown in Figure 8.3. It depicts the main groups of data elements that are necessary for UDC maintenance, distribution and use.

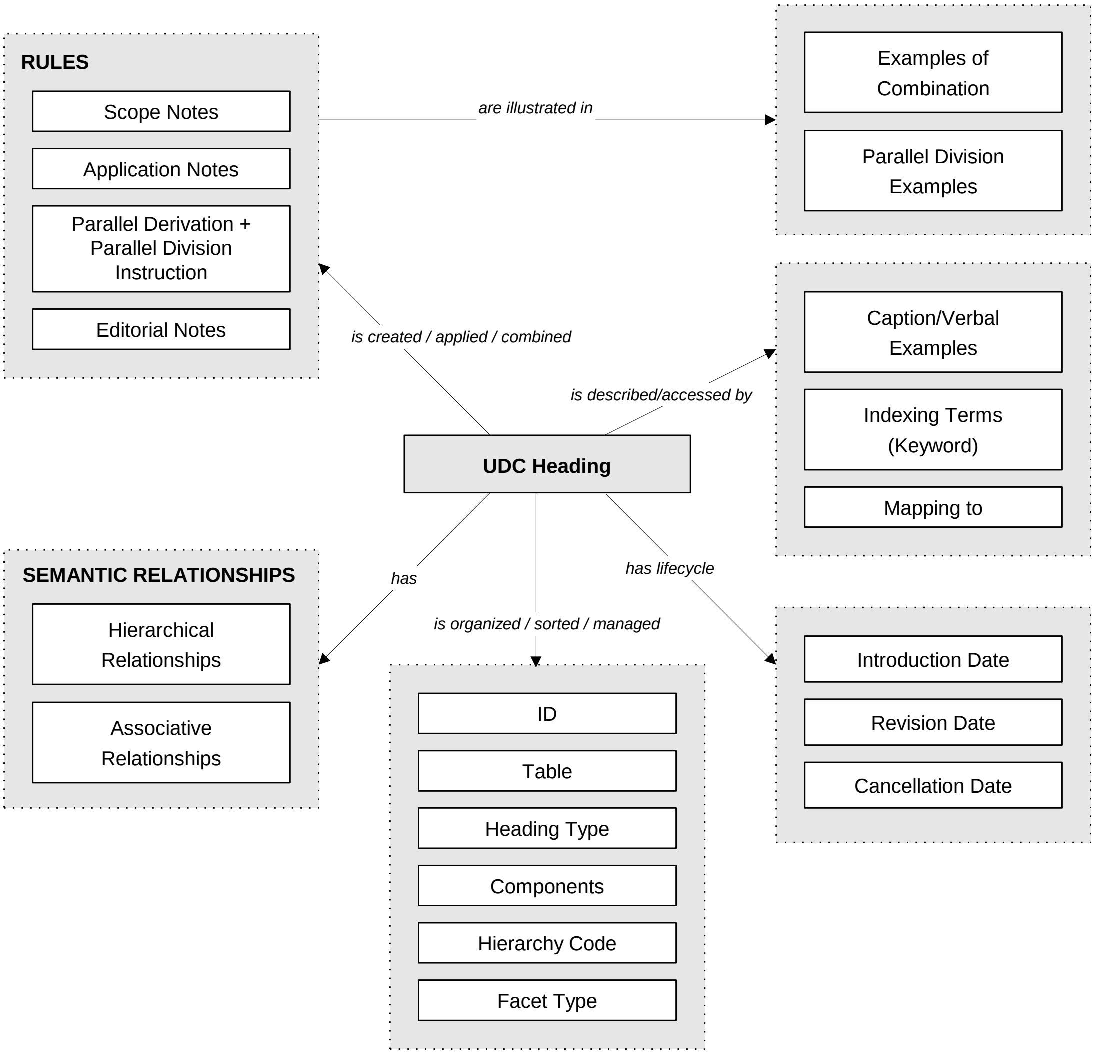


Figure 8.3: UDC conceptual schema

This conceptual model closely examines the group of data necessary in the management and maintenance of UDC headings and lists all the information that might improve the management and application of UDC as mentioned earlier.[153]

ID stands for a UDC *unique identifier*, which is numeric and automatically generated, and ensures the permanent identification of every UDC concept. If a UDC number is cancelled and re-introduced with a different meaning (which cannot be completely ruled out in the revision process), its *ID* could be used to distinguish between the two concepts. This is essential in managing updates and in the exchange of data. *Components* represent a structured and coded breakdown of the UDC heading into pre-composed elements, when this is a pre-synthesised type of notation. When this information is available at the actual source of UDC data, it may greatly improve and encourage better management of pre-synthesised numbers in the target system. The *Hierarchy code* is purely numeric data that holds the exact hierarchical and filing positions of each notation in relation to the system as a whole and this can be automatically generated, based on data from the broader class (hierarchical relationship) and stored in a separate table in any relational database. These data, if distributed with UDC, can help in hierarchy management, display and the control of UDC sorting. *Facet type* (entity, kinds, parts, processes, operations, agents etc.) is coded data that supports the maintenance and further development of a vocabulary based on facet analysis, if such an approach were planned.

When one engages in the creation of a specific tool, as was the case with the *UDC Research database*, created for the purpose of this research, a more precise representation of data elements is necessary. This can be achieved easily following the conceptual model. For this purpose one can choose an entity relationship model approach. Here, instead of showing data entity groups and data flow, the UDC data model is broken down into isolated components (entities), their attributes and associated relationships.

The entity relationship model shown in the following, Figure 8.4 is created based on experiences during manual validation of UDC data. One of the main objectives of this model was to avoid the pitfalls in UDC notation validation. Another objective was to provide all the elements required for further improving and

---

[153] Some of the data mentioned do not exist in the UDC MRF and are not even proposed to be added in the new database format (hierarchy code, facet type, unique UDC heading identifier, edition type, mapping to). These data, however, are often used in library systems or different kinds of IR systems or classification tools.

enriching UDC data and ensuring its straightforward application in an online environment.

This database format can be also created with any database software including CDS/ISIS. The reason why a relational database (MySQL) was chosen to test this conceptual schema was the easy implementation of automatic validation of every simple notational element entered anywhere in the UDC record against the existing UDC MRF records. In addition, SQL statements created in this database can be exported and re-used in other relational database management system which makes the solution easy to re-use and implement in different applications.

As shown in the UDC Heading box in Figure 8.4, each *UDC Heading* may have a number of attributes that may occur only once. These attributes are separated from entities that have multiple occurrences and have to be controlled through separate relational tables. In this model (representing a multilingual database) all the entities and attributes that may have an associated language are maintained in a separate table (caption, caption verbal examples, scope notes, application notes, editorial notes). Also, in order to keep each *example of combination* and *parallel division example* associated with its description, these had to be modelled as separate entities. In addition, based on problems experienced with notation validation, it was decided that all entities and attributes that could contain pre-synthesised notations and ought to be traced to special auxiliary tables or exist as a result of parallel derivations - are linked to a *notation validation* table (such entities/attributes are scope notes, application notes, examples of combinations and parallel division examples).

This database model assumes that complete UDC data from 1993 onwards would be maintained and managed throughout its lifecycle. Cancelled numbers would be 'flagged' as deleted but would stay in the database. Each new UDC heading would be associated, if applicable, with the heading it replaced, and likewise, all cancelled headings would have links to the headings they are replacing (if applicable). The existence of an independent unique identifier would allow automatic update in target systems and linking between 'old' and 'new' UDC editions. 'Edition type' (e.g. extended, abridged or MRF standard) would enable different editions to be kept together and crosschecked. Each UDC heading could be managed for different editions, depending on the business model in place.

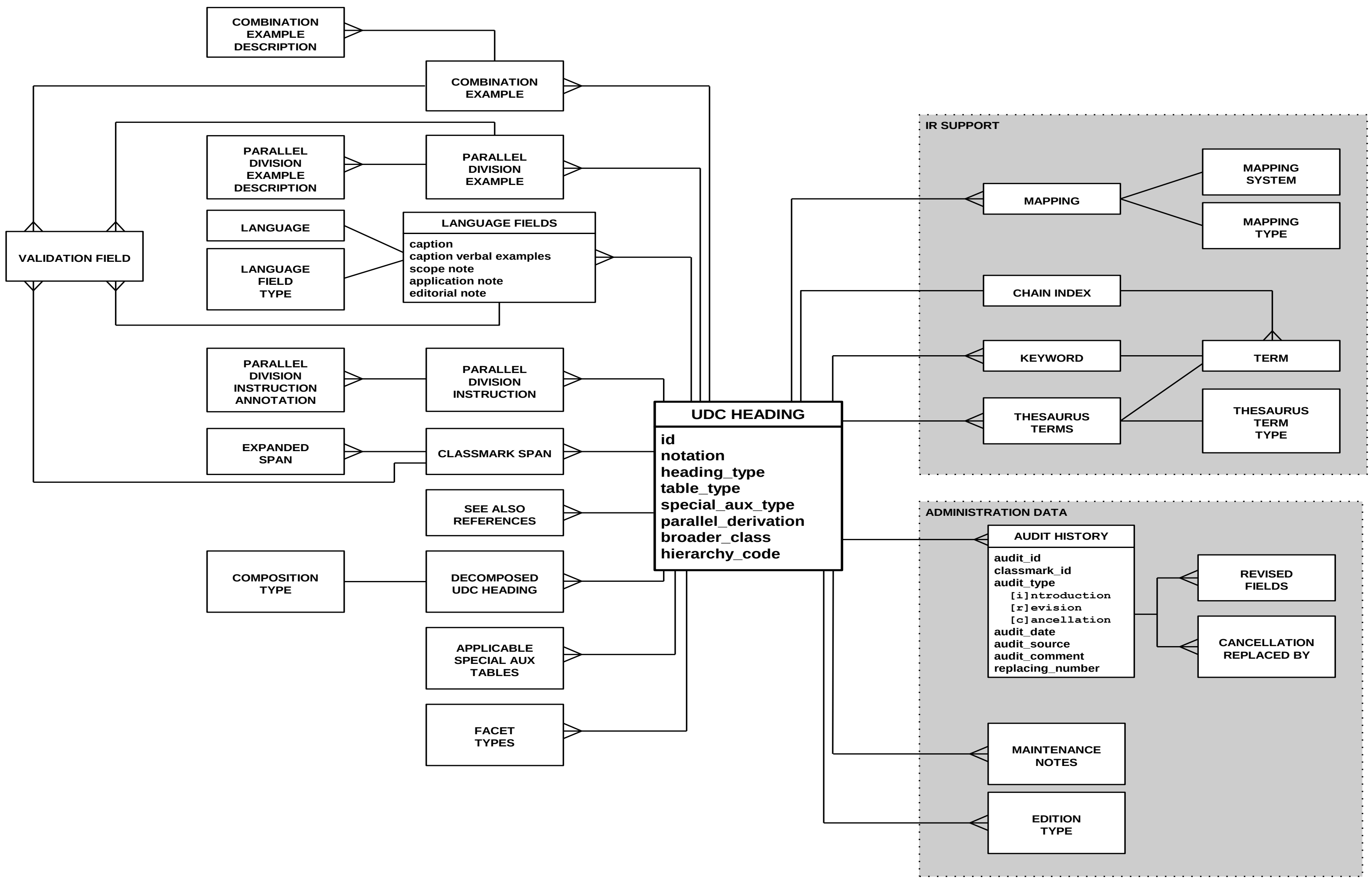


Figure 8.4: Entity relationship model of the *UDC Research Database*

In terms of UDC system enrichment, this data model supports the incremental creation of verbal access points either through keywords, mappings to other systems or thesauri. Based on this entity relationship model, a detailed database element schema was created (shown in Appendix 10.3.). As is evident from this schema, this would enable tools to be created that would enable automatic initial population and manual editing of keywords, subject-alphabetical indexes or thesauri. Generally, this model assumes that the database tool could be used to support different services. Overlaying this data structure, a more versatile business model would determine the number of applications and tools that could be built upon it: update tools, product distribution tools, vocabulary development tools, translation tools, classification tools or terminology service tools. Vocabulary development tools and translation tools based on this data model could be made available online for all temporary work to be uploaded and collated. This would be a much more efficient way for contributors to upload their proposals, which could then be made available for peer review and validation against the existing system, well in advance of the planned final edition.

The purpose of thinking through the UDC data model, modelling and element schema, is to illustrate how the direction of development might produce database tools that are more suitable for both UDC maintenance and its exchange and use in a networked environment. This research shows that such a database tool, which greatly increases the usability of UDC and can further improve UDC vocabulary, is relatively simple to create. The improvement and enrichment of UDC data at source is likely to influence all online use that is based on the UDC MRF and is likely to make UDC a more desirable indexing system. Better quality UDC source data means a decrease in costs, which are currently deferred to users and implementers. The reach of the proposed changes depends, however, on the variety of different data exports and their flexibility with respect to vocabulary exchange standards and relevant vocabulary transport formats. Because the UDC is the second most used classification system in the world, its data model could be exploited to influence standards development.

These proposed changes for the UDC model and maintenance database are concerned only with improvements achievable on the data and data management level which can influence the value of the UDC in the networked environment. These improvements cannot, however, reduce the number of problems that may arise from any deficiency in the area of UDC management, maintenance and distribution, indicated earlier in this research.[154] The addition of new data would require an elaborate and updated business model, a detailed projection of resources to be put into the database as part of a basic business infrastructure and planning over a period of time.

8.6 Summary of UDC data and data model

The preceding four chapters discussed how the UDC can and should be used in information retrieval (Chapter Four) and how it should be interfaced and what level of interface functionality there should be in a real life environment (Chapter Five). Obstacles to the application of the classification online were identified as a lack of proper data format to hold structured classification data and a lack of standards that would serve as a vehicle for its exchange and use in the networked environment (Chapter Six). Then 'real-world' usage, users, the environment and conditions in which UDC is maintained, managed and distributed were examined in Chapter Seven. Finally the UDC data source was analysed with respect to all functions and applications that may be built upon it.

This research indicates that the form and formats in which the UDC is distributed fall short of actual users' needs in a predominantly online environment. But, most importantly, the requisite improvements seem to be relatively easy to achieve. For instance, the quality of UDC data does not seem to be sufficiently controlled or managed and data validation shows that improvements can be made even with existing database tools. In addition, although the new database format proposed by Riesthuis may not correct problems in quality control management, it is likely to increase the amount of data necessary for online applications significantly. This is particularly so with respect to providing data for linking every main UDC number to its respective special auxiliary table. Also, by providing data on broader classes, hierarchical browsing and search expansion can be implemented. Finally, the proposed addition of index terms to each class may help in the creation of a subject-alphabetical index to the classification. The new database format proposal, however,

---

[154] Accessibility and affordability of the UDC data to existing and potential users was certainly an important issue during the period 1993-2004. It should be noted that data was considered expensive even in the period in which the Consortium put no funding into the database, UDC data export or data quality improvements (cf. Wagner, 2000).

does not address all issues relevant to the use of the UDC online.

This research proposed a UDC conceptual schema that shows possible ways in which a classification system can be presented, expanded and made more suitable for machine-processing. This is further illustrated with the example of a UDC tool entity relationship model and a data element schema.

Observations on the management and maintenance of the UDC throughout this research were based on information available to any user of the system: the database itself, accompanying documentation, published articles and the UDC Consortium communications to users via the *UDC Extensions and Corrections to the UDC*. If the period after the initial automation of the schedules to 2005 is taken as an indicator, further improvement of the UDC data format and the database tool was not a priority for the classification owner. The potential significance and strength of the UDC system in the networked environment and the ease with which this potential can be technically achieved ought to be, therefore, assessed independently of the Consortium management policy at the time of this report.

CHAPTER NINE: CONCLUSION

The original assumption at the beginning of this research was that, from the 1990s onwards, the way in which the UDC was used had changed. One reason for this change was the fact that the UDC was stored and distributed in database format, which facilitated developments for its use online. Other reasons included the creation of an open networked environment. With an increased need for known and recognizable knowledge organization tools on the Internet, the classification has become more interesting to a number of new users outside the bibliographic domain. The Internet has enabled the creation of a highly collaborative information environment where the practice of sharing and exchanging classifications and other controlled vocabularies is common. Classification sharing in this environment depends upon accessing and processing the same data by different distributed systems, which reduces duplication of effort and leads to better system-interoperability.

The starting point of this research was that the use of KOS, such as classifications, would depend upon data being machine-readable and also available in a standardized format. After an initial review of the literature concerning the trends and developments in information retrieval (especially developments on the Internet with respect to resource discovery), it became clear that developments in the Internet information space would ultimately determine the future of any specific indexing language, irrespective of its provenance and use in the past.

Trends in information retrieval systems in the 1990s moved towards employing interactive and flexible information retrieval techniques that combined searching and browsing. This was advantageous to classifications as they can support both approaches in information seeking. In addition, the advantage of using classification for subject browsing became obvious and justifiable with the development of the GUI and Web-based interface functionalities.

Towards the end of the 1990s, traditional library catalogues started migrating to graphically enhanced and more powerful Web OPACs that often bridged several collections and/or several library systems. Windows-based operating systems, hypertext, and most of all, Internet technologies, formed new information-seeking habits. In spite of the fact that libraries have remained, more or less, uninterested in the automation of subject indexing languages and their greater exploitation in information retrieval, significant changes have occurred:

- the extensive use of knowledge organization tools outside the library domain
- the publishing of classification schemes on the Internet
- the application of library classifications for the (automatic) classification of Internet resources
- the creation of classification authority formats to hold and maintain classification schemes
- the creation of independent standards and formats for the exchange of knowledge organization tools
- wider interest in the automation of the classification process (automatic indexing)

This clearly indicates that the exchange of information in a networked environment relies on a high level of computer readability of knowledge organization systems. Therefore any use of the UDC in this environment will depend on its availability in a machine-readable format that can be shared between different systems and platforms.

This research has confirmed the initial assumptions that the way in which the UDC is used in information systems, particularly subject gateways and library catalogues, does not exploit the full potential of the classification for supporting subject access. It also became evident that the inadequate machine-readability of the classification impedes the ease of implementing the system online and thus increases its cost of use.

The research has proved that UDC is suitable for knowledge organization in the global network and has explored its values in information retrieval that justify efforts for its implementation. The requirements for UDC automation and its full exploitation and exchange in the current information environment have been summarised. And finally the improvement of the UDC database format and a suggestion for a more robust data model and data element schema were proposed that would facilitate and contribute to the use of the UDC system in a networked environment.

9.1 Findings

The quality of the system. The combination of the size of the UDC vocabulary and a powerful analytico-synthetic syntax makes the classification a strong candidate when there is a need for:

i) hierarchical organization of information resources

ii) combination of browsing and searching of information resources and

ii) scalable subject indexing (i.e. general/broad and very exhaustive/specific indexing)

The advantage of UDC's semi-faceted, analytico-synthetic structure is that it is less rigid when it comes to the underlying discipline-based concept organization, as it allows for great flexibility in relating disciplines and subjects, thus expressing interdisciplinary and multidisciplinary relationships and allowing for easier subject expansion and accommodation of new knowledge. Thus the UDC's power in indexing is even greater than could be anticipated from the already significant size of its vocabulary (66,737 classification numbers in UDC MRF 2004). Its strength in IR is that it can be used to support subject browsing (hierarchical and associative), pre-coordinated indexing and post-coordinated searching and search expansion.

The automation of the UDC schedules in 1993, which at that time had been reduced to a standard size of around 60,000 numbers, has helped updates and distribution become faster and more efficient. During the period 1993-2005 the UDC schedules have undergone constant revision and improvement, and in this period there were no fewer than twelve annual updates of the UDC released and distributed to users.

In 2005, the UDC is the second most widely used classification system in libraries and information centres, despite a general decline in the popularity of classification within libraries in the eighties and nineties. As has been demonstrated, UDC is used in at least 112 countries across all continents and in 39 languages. Although it is still primarily used in the bibliographic domain, metadata standards developed for information resource discovery on the Internet regard UDC as a 'standard controlled vocabulary' and recommend its use in the subject metadata of Internet resources.

In the period 1993-2004, UDC was considered a candidate for many networked information services. The use of the UDC during this period in nine information gateways (English language), two of which were based on automatic indexing, and an increase in the number of information gateways using UDC in Eastern European countries after the year 2000, illustrates the interest expressed in the UDC by the Internet community.

Only a few other general classifications share the UDC's 'high international profile', primarily DDC and LCC. But unlike DDC and LCC, the UDC has been made available and distributed to users as a database file. In a networked environment, where tools for indexing and retrieval are usually shared between distributed systems, this is an important advantage. It allows users to customise UDC data and implement it directly into their own systems, while continuing to enrich the data in order to improve subject access points.

Requirements for UDC use in networked environments. An examination of UDC use in IR confirmed that its successful implementation was closely tied to the possibility of processing individual elements of classification data automatically. The full advantage of using UDC for browsing and searching depends on the possibility of managing its semantic (hierarchical and associative) and its notational pre-composed structure, and also on the possibility of linking notation to verbal expressions.

The development of a classification authority file that performs this data management function is very important for UDC, since it is an analytico-synthetic classification with classification headings expanded and recombined in the process of indexing. Thus, the classification notation, when implemented in an information retrieval system, has added value compared to the classification source vocabulary. In this context an authority file serves as a tool for recording, re-using, sharing and searching customised/customisable classification content. Observations of thirty OPACs using UDC showed an obvious trend towards the creation of classification authority tools in order to improve subject access in both in-house and vendor library systems.

UDC data improvement. The UDC data format has to support two distinct but related functional levels: *(a)* UDC system management and maintenance and *(b)* UDC data use and sharing. The examination of the UDC MRF data indicated that the present data format, while currently distributed as proprietary data, lacks a significant amount of machine-readable information, detailed below, that is necessary for the management of UDC and its application in an online environment:

- coding components of pre-synthesised notation
- coding hierarchies independently of notation
- a unique UDC number identifier
- verbal access to notation (in addition to the caption)

If UDC source data contained fully coded pre-synthesised notations, searching UDC would be more efficient and easier to implement in individual systems. Similarly, information on class hierarchies is essential for filing and browsing and should be maintained and distributed as a standard part of UDC source data, since UDC notation is not reliable in expressing hierarchy. The lack of a unique UDC number identifier impedes the automatic control of classification authority file updates and is vital for classification exchange. The omission of controlled verbal access to UDC notation increases the cost of its implementation in IR.

At the present time (2005) the creation of an authority file based on UDC MRF data requires a great deal of data customisation, which is both a duplication of effort by individual users and a waste of their resources. There is, therefore, a strong case for making improvements in the following areas:

- facilitating classification authority file creation by providing more useful classification source data which would reduce the duplication of user-effort
- providing exports in standard data formats that would ease UDC implementation in libraries and other IR systems
- active participation in the improvement of standards for vocabulary exchange that could serve as a vehicle for the sharing of existing classification authority data within a community of users

The logical requirement would be for the UDC MRF to be distributed in a format that would be usable in library systems without imposing significant additional costs for data conversion. In terms of data format and transport (data portability), there are at present two standards that have been created with classification authority control in mind, specifically: MARC 21 and UNIMARC classification authority formats. Both these formats would, in theory, satisfy perfectly the function of classification management and exchange in a library community, as they would be supported by vendor library systems.

Closer examination of these two formats, however, indicates that they do not provide the means for managing notational components and are therefore not suitable for analytico-synthetic classifications. But UNIMARC is used in countries using UDC, is still in development and is a good candidate for becoming a suitable format for the exchange of analytico-synthetic classifications. Proposals for its improvement were suggested as a part of this research.

This study indicated that the prerequisite for any improvements in UDC transport and exchange using standard formats is first to improve the UDC MRF data format itself. The reason for this is that the present UDC MRF data format is insufficient for supporting the basic functions of maintenance and validation of UDC data that underpins the entire maintenance and distribution business. It is not fully adequate for the creation of online tools or for straightforward application in IR systems. In addition, if it were complete and functional, the UDC MRF data format could serve as a model of good practice in processing UDC, which would then be more likely to influence standards for classification exchange.

The study of UDC data quality, its data structure and data model shows that the need for improvement of UDC data is obvious and has been evident ever since the first database was created in 1993. The requirements for data enrichment are clear and straightforward. The improvements required of the data format, which would further enhance the value of UDC in a networked environment and would be easy to implement have been set out in the foregoing chapters. The improvements of UDC data source, however, ought to be followed by the creation of better, more versatile export formats, some of which should be based on existing standards for vocabulary exchange such as MARC21, UNIMARC or XTM.

9.1.1 Explanation of findings

The main change with respect to the use of classification in the nineties was the realisation of the relative value of knowledge organization tools. Under financial pressure, libraries in many countries, especially in the English-speaking world, continued the previous decade's trend of reducing cataloguing and indexing costs and purchasing bibliographic packages such as the OCLC bibliographic database. Bibliographic records in this database already contained DDC and LCC numbers and LCSH. For many libraries this meant that they would not normally need any cataloguing and classification providing they decided to use the suggested controlled vocabularies. For this reason, some libraries and bibliographic services in Europe changed from using UDC or other classification systems to using DDC. It is evident that the choice of classification system in the 1990s continued to be influenced by the costs the system incurred to library services.

The size and heterogeneity of the Internet information space increased the need for KO tools such as classifications but their value came under scrutiny in the areas of ease-of-use, low implementation cost and machine-readability. A section of the KO community dealing with Internet resource discovery in non-profit information services expressed concern at the cost of classifications and their amenability to machine processing when evaluated against the intellectual quality of the classification system structure. The reason for this was that resource discovery may benefit instantly from almost any kind of classificatory structure combined with other techniques of automatic or semi-automatic resource indexing, provided this semantic structure can be implemented and put to use at low cost with no need for additional expert human resources. A classification system that cannot be implemented online without significant additional costs (human resources and software), in spite of its intellectual excellence, is not likely to be adopted and shared by a great number of information providers and therefore is not likely to become widely used on the Internet.

In principle, the Internet information space is not biased towards any kind of KO tool or any specific indexing language. Such a tool, however, is likely to be chosen on the basis of being ready for implementation. In addition, advantage is always given to indexing languages that can be mapped on to other indexing languages and thus provide multiple subject access, which is more suitable for linking distributed information services.

In the environment of the global information space, human labour is expensive, has limited reach and ought to be used effectively. Information exchange on the Internet is based on services that are increasingly moving towards m2m applications. Human indexing is the most valuable and expensive activity and has to be more efficient and supported with automated and semi-automated tools that can save time, reduce the occurrence of mistakes and increase the re-usability of the labour invested.

Developments linked to Semantic Web technology, which became the main focus of the Internet community at the end of 1990s, have further promoted the importance of metadata infrastructure and controlled vocabularies in resource discovery while stressing the importance of m2m applications. The rôle of classification with respect to these developments is illustrated well in Chapter Three, observing the use of UDC on the Internet. There was no increase in the initial interest in exploring the use of UDC in an information gateway in the period 1993-2004. When observed closely, the reason for this seemed to be a lack of free general information on UDC, especially documentation on system structure and online functioning, a lack of freely available information on the UDC data format, and a lack of XML exports. Most of all, it was the implementation cost combined with both the cost of a UDC MRF licence and the long and unpredictable waiting time required to acquire information on a data licence.

It is obvious that Internet users thoroughly question the cost of implementation in an m2m environment and the quality and availability of classification data. The Internet community is very interested in systems whose global use for resource discovery is not impeded by a proprietary format, the need for too much human input, localised use, or systems crippled by unmanageable copyright restrictions that are slow to negotiate.

UDC translation into a significant number of languages enhances its prospects for global use. For a networked application, the availability of schedules in more than one language in a machine-readable format is very important. The existing total of twelve electronic editions, only three of which are multilingual, shows however, that in spite of the fact that the UDC MRF has been available as a database file since 1993, the publishing of online editions is not easy or cost effective and involves both duplication of effort and financial investment by individual licence purchasers. The Consortium policy towards multilingual editions is to charge for UDC MRF data and then pass the issue of copyright agreements and further costs over to the publisher, who owns the publishing rights for individual languages. This does not seem to encourage or in any way facilitate the production of multilingual editions and makes the undertaking expensive and slow, involving a great deal of duplicated work. As a result, each UDC online publication is a project in its own right, which has to develop its own software application with a specific data structure and coding.

The improvement of the original data format in a way that can save the cost of UDC translation, publishing online, implementation and use would not only be paramount for users but also be urgent for appropriate UDC data maintenance. The new improved database format proposed in 2003 by Riesthuis suggested improvements that are well overdue and are most urgent with respect to the above-mentioned problems. In addition, there would still be room for further development and enrichment of UDC data in terms of mapping and verbal access points, and the provision of customised subject, language or size orientated packages of UDC data for different purposes. With the development of the Semantic Web, it would be wise to accommodate the possibility of further improvement of the machine-readability of UDC data.

This research was focused on the quality of UDC data and their improvement. It illustrated that an improved implementation of the classification and better use can be achieved easily through enrichment of the UDC data source. But these issues have to be examined separately from the way the UDC is managed and maintained and with respect to the business goals of the current owner of the UDC. Should the UDC owner decide to change its business policy to be more in-line with the current use of the system, the UDC data and the way it is maintained and distributed would require change. Discourse with users and transparency of product content, data format, practical value and price should be addressed, but enhanced and more varied export formats are the most urgent issue.

The quality of UDC content and its global usage is a valuable asset with great potential still largely unexploited in the networked environment. The investment put into its modernisation and vocabulary expansion by the Consortium during the period 1993-2004 is yet to be fully appreciated by users and this can be achieved through forward looking, application orientated and standardized data exports.

9.2 Research limitations

This research was primarily interested in the qualities of the UDC system for information exchange in a networked environment and how the available UDC MRF data correspond to the needs of users and applications in this environment. In retrospect, it can be said that its scope and final outcome should be viewed in relation to some objective limitations.

In looking into the use of UDC on the Internet and presenting its full spread in Internet information services, there was an obstacle in evaluating the presence and use of UDC outside Western and Central Europe. It was discovered that different scripts and languages presented a significant limitation to this specific study. When used in supporting resource discovery services, UDC is usually not present on a gateway interface and until documentation is available on the system architecture and indexing, it is not possible to establish the type of application. Seeking out and analysing this documentation would require more time and more extensive study.

Some improvement of the existing UDC data format has been proposed in order to make the scheme more suitable for the networked environment and its straightforward online implementation, and it has been demonstrated that this is technically easy to achieve. One of the recommendations is that there is a need to make the UDC MRF data available in some of the standard formats for vocabulary exchange. MARC formats would be, for instance, a natural choice for the bibliographic domain while XTM (XML schema for Topic Maps) seems to be the logical choice between domain non-specific and network orientated vocabulary standards. The next step is mapping the present UDC MRF data format to some existing vocabulary exchange standards.

The reason why this cannot be done yet is that the present UDC MRF data format is lacking various important data elements, without which the quality of a proposed mapping would be poor and would hardly justify the effort. Because of the uncertainty of whether the UDC MRF format will actually be improved, based on the Riesthuis' format change proposal from 2003, it also seems pointless to work on data mapping with respect to this specific yet unapproved format.

9.3 Further research

The straightforward practical application of this research would result in the improvement of the UDC data format and UDC data exports, which would then open a plethora of issues that are waiting to be examined. One area of research is the community of UDC users, their needs and their systems' requirements for machine-readability of classification data. Another area awaits further research into vocabulary standards and exchange in the networked environment.

The first research area, UDC practice and the UDC user community, calls for different types of study. In the course of this research it was felt that further examination would be needed into the actual management of classification data in library systems and metadata repositories with respect to the actual system architecture, information discovery across distributed systems and subject access control. A more focused examination of the way the UDC MRF is actually used by libraries would be very important and an insight into the different practices in UDC authority control and classification authority data exchange in the bibliographic domain would be most valuable. This would help create an awareness of the critical mass of users that depend upon the improvement of MARC classification formats and users' real needs with respect to UDC MRF data.

The study of the interface functionalities of UDC OPACs has only touched on the important issue of the rôle of classification in IR. A further carefully designed user study involving good quality OPACs would be necessary to determine the value of the UDC to users and their ability to make use of the classification (hidden or exposed) in browsing and navigating a subject area. Similarly, research would be necessary to examine the value of semantic relationships established in the UDC in creating a browsing taxonomy in subject gateways.

Finally, it would be advantageous to undertake further studies into the actual UDC user community worldwide. The study conducted here established very basic information, showing that UDC users in 112 countries can be contacted by e-mail. For the majority of these users, their library catalogue can be accessed on the Web. An extended study would be necessary to establish actual usage in these countries and the size and types of collection indexed by the UDC. In addition, it would be interesting to establish the situation in the 55 countries from which no data were collected.

The second area of research leads towards standards for vocabulary exchange in the networked environment. An in-depth comparative study into standard formats for exchange of subject vocabulary data in the networked environment would be required. The portability of UDC data can be examined and tested through the various new standards for networked vocabulary exchange that already exist or are being created in the course of this research (e.g. SKOS, XFML, or VDEX). This would help establish which data formats and data transport standards, apart from MARC formats, would be the most suitable for classification and for what purposes in information discovery. One exciting new area of study would be the extension of existing research on the proposed UDC data model and MARC formats into a generic classification format or *meta-classification schema.*

When it comes to UDC data sharing in the networked environment, *terminological services* are the next most important model to explore. Such a development requires more technologically orientated research involving real user test beds and mapping systems. This is, however, entirely dependent on the Consortium's business model, whose affinity for expansion and change with respect to the networked environment would require further examination.

The fact that UDC data exist and can already be purchased in a structured data format opens many possibilities in researching the limitation and potential of classification data in general or the format of data itself. The wealth of UDC vocabulary, if more accessible and affordable to researchers, would be welcome and would probably revive interest in the UDC, thus opening many as yet unexplored possibilities for its use.

BIBLIOGRAPHY

ABOUD, M. et al. (1993) "Querying a hypertext information retrieval system by the use of classification", *Information Processing & Management*, 29 (3) 1993, 387-396.

ADLER, E. (2000) "Multilingual and multiscript subject access : the case of Israel", *66th IFLA Council and General Conference, Jerusalem, Israel, 13-18 August 2000*. http://www.ifla.org/IV/ifla66/papers/035-130e.htm.

AGOSTI, M. (1991) "Usage of current hypertext systems as powerful information retrieval", *Documentary languages and databases : papers from the Rome Conference, Dec 3-4, 1990*. Edited by G. Negrini, T. Farnesi, D. Benediktsson. Frankfurt/Main : Indeks Verlag, 1991. (Advances in knowledge organization 3), 213-119.

AITCHISON, J. (1970) "The thesaurofacet : a multipurpose retrieval language tool", *Journal of Documentation*, 26 (3) 1970, 187-203.

AITCHISON, J.; GILCHRIST, A.; BAWDEN, D. (2000) Thesaurus construction and use : a practical manual. 4th ed. London : Aslib, 2000.

ALLEN, B. L. (1998) "Visualisation and cognitive abilities", *Visualizing subject access for 21st century information resources : proceedings of 34th Annual Clinic on Library Applications of Data Processing, Champaign, 1997*. Edited by P. Atherton Cochrane, E. H. Johnson. Champaign, IL : Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 1998. 66-79.

ALLEN, R. B. (1995) "Two digital library interfaces that exploit hierarchical structure.", *DAGS95, Electronic Publishing in the Information Superhighway, Boston, Massachusetts, May 30 - June 2, 1995*.
http://www.raven.umd.edu/~rba/PAPERS/LIBR/libr.html.

ALLEN, R. B. (1996) "Retrieval from facet spaces", 1996.

http://raven.umd.edu/~rba/PAPERS/FACETS/facets.html.

ANAGNOSTELIS, B.; COOKE, A.; McNAB, A. (1997) "Thinking critically about information on the Web", *Vine*, 104 1997, 21-24.

ANDERSON, J. D. (1989) "Indexing and classification : file organization and display for information retrieval", *Indexing : the state of our knowledge and the state of our ignorance : proceedings of the 20th Annual Meeting of the American Society of Indexers*. Edited by B. H. Weinberg. Medford, NJ : Learned Information, 1989, 71-83.

ANDREASEN, T.; SCHOMACKER, T. (1997) "DanBib : a union catalogue applied for user friendly flexible querying", *63rd IFLA [Council and] General Conference, Copenhagen, Denmark, August 31- September 5, 1997*. http://www.ifla.org/IV/ifla63/63andt.htm.

ANDREW, D.; NONNECKE, B.; PREECE, J. (2003) "Conducting research on the Internet : online survey design, development and implementation guidelines". http://www.ifsm.umbc.edu/~preece/Papers/Online_survey_design_IJHCI04.pdf.

ANJER, J. (2004) UDC in Norway". 2004-08-27. Private e-mail message.

APTAGIRI, D. V; GOPINATH, M. A; PRASAD, A. R. D. (1995) "A frame-based knowledge representation paradigm for automating POPSI", *Knowledge Organization,* 22 (3/4), 1995, 162-167.

ARDÖ, A.; KOCH, T. (1999) "Creation and automatic classification of a robot generated subject index", *Digital Libraries '99, The Fourth ACM Conference on Digital Libraries, August 11-14, 1999 in Berkeley, California*. http://www.lub.lu.se/desire/poster.htm.

ARDÖ, A; KOCH, T. (1999a) "Automatic classification applied to full text Internet documents in a robot-generated subject index", *23rd International Online Information Meeting, London, 7-9 December 1999 : proceedings.* Edited by B.

McKenna. Oxford : Learned Information Europe Ltd., 1999, 239-246.

ARDÖ, A. et al. (1994) "Improving resource discovery and retrieval on the Internet : the Nordic WAIS/World Wide Web Project : summary report". NORDINFO project, 1994. http://www.lib.lu.se/W4/summary.html.

ARDÖ, A. et al. (2000) "Browsing engineering resources on the Web : a general knowledge organization scheme (Dewey) vs. a special scheme (EI)", *Dynamism and stability of knowledge organization : proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada*. Edited by C. Beghtol, L. C. Howarth, N. J. Williamson. Würzburg : Ergon Verlag, 2000. (Advances in knowledge organization 7), 385-390.

ARMSTRONG, C. (1997) "Metadata, PICS and quality", *Ariadne* (The Web Version), May 1997. http://www.ariadne.ac.uk/issue9/pics/.

ARNOLD, S. E. (1999) "Search engines 2000 : rough running, prone to stall", *23rd International Online Information Meeting, London, 7-9 December 1999 : proceedings*. Edited by B. McKenna. Oxford : Learned Information Europe Ltd., 1999, 97-103.

ARSKIY, Yu. A.; NESTEROV, O. P. (2000) "VINITI : The All-Russian Institute for Scientific and Technical Information", *Extensions & Corrections to the UDC,* 22 2000, 7-8.

ASHTON, S.; ROBERTSON, S. (2000) "Re-casting our net : broadening information access at the National Maritime Museum", Museums and the Web 2000, April 16-19, 2000, Minneapolis, Minnesota, USA.
http://www.archimuse.com/mw2000/papers/robertson/robertson.html.

AUD-FRANKLIN, R. (2003) "Re-inventing subject access for the Semantic Web", *Online Information Review,* 27 (2) 2003, 94-101.

AUSTIN, D. (1978) "A conceptual approach to the organization of machine-held files for subject retrieval", *Conceptual basis of the classification of knowledge : proceedings of the Ottawa Conference, October 1st to 5th 1971*. Edited by J. A. Wojciechowski. New York; München; Paris : K.G. Saur, 1978, 371-398.

AUSTIN, D. (1979) "Differences between library classification and machine/based subject retrieval systems : some inferences drawn from research in Britain, 1963 - 1973", *Ordering systems for global information networks : proceedings of the Third International Study Conference on Classification Research held at Bombay, India, during 6-11 January 1975*. Edited by A. Neelameghan. Bangalore : DRTC : FID/CR : Sarada Ranganathan Endowment for Library Science, 1979. (FID 533), 326-340.

AUSTIN, D. (1982) "Basic concept classes and primitive relationships", *Universal classification : subject analysis and ordering systems : proceeding of the 4th International Study Conference on Classification Research, 6th Annual Conference of Gesellshaft für Klassifikation, Augsburg, 28 June - 2 July 1982, Vol. 1*. Edited by I. Dahlberg. Frankfurt : Indeks Verlag, 1982, 86-94.

BACSICH, P. et al. (1999) "The standards fora for online education", *D-lib Magazine*, 5 (12) 1999. http://www.dlib.org/dlib/december99/12miller.html.

BAKER, T. (1997) "Metadata semantics shared across languages : Dublin Cores in languages other than English : report", *Fourth Dublin Core Workshop in Canberra, Australia, 3-5 March 1997*.
http://dublincore.org/groups/languages/mr-19970303.shtml.

BAKER, T. (2000) "A Grammar of Dublin Core", *D-lib Magazine*, 6 (10) 2000. http://www.dlib.org/dlib/october00/baker/10baker.html.

BAKER, T. et al. (2001) "What Terms Does Your Metadata Use? : application profiles as machine-understandable narratives", *Journal of Digital Information*, 2 (2) 2001. http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Baker/.

BALIKOVÁ, M. (2001) "Application of UDC in Czechia", *Extensions & Corrections to the UDC*, 23 2001, 12-14.

BALIKOVÁ, M. (2003) "Subject indexing and classification development in Czechia in 2002-2003", *Extensions & Corrections to the UDC,* 25 2003, 9-10.

BALIKOVÁ, M; STOKLASOVÁ, B.; PSOHLAVEC, S. (1998) "Bilingual (English-Czech) UDC Master Reference File on CD-ROM", *64th IFLA [Council and] General Conference, Amsterdam, August 16 - August 21, 1998.* http://www.ifla.org/IV/ifla64/104-158e.htm.

BAMPTON, R.; COWTON, C. J. (2002) "The E-Interview", *Forum Qualitative Social Research,* 3 (2) 2002. http://www.qualitative-research.net/fqs-texte/2-02/2-02bamptoncowton-e.htm.

BARNHART, L. (1996) "Access control records : prospects and challenges", *Authority Control in the 21st Century : an invitational conference, Dublin, OH, March 31 - April 1, 1996.* http://digitalarchive.oclc.org/da/ViewObject.jsp?objid=0000003520&reqid=354.

BATES, M. (1989) "The design of browsing and berrypicking techniques for the online search interface", *Online Review,* 13 (5) 1989, 407-431. http://www.gseis.ucla.edu/faculty/bates/berrypicking.html.

BATTY, D. C. (1970) "Chain indexing", *Encyclopedia of library and information science.* Edited by A. Kent, H. Lancour. New York : Marcel Dekker, 1970, 4, 423-434.

BATTY, D. C. (1981) "Universal Decimal Classification", *Encyclopedia of library and information science*. Edited by A. Kent, H. Lancour, J. E. Daily. New York : Marcel Dekker, 1981, 32, 141-146.

BEALL, J. (1992) "Editing the Dewey Decimal Classification online : the evolution of the DDC database", *Classification research for knowledge representation and organization : proceedings of the 5th International Study Conference on Classification Research, Toronto, Canada, 24-28 June 1991*. Edited by N. J. Williamson, M. Hudon. Amsterdam : Elsevier Science Publishers; The Hague : FID,

1992. 29-37.

BEALL, J. (1996) "Dewey for Windows", *62nd IFLA General Conference, 25-31 August 1996, Beijing, Workshop on Dewey Decimal Classification : edition 21 and international perspectives* (paper handed out at the workshop).

BEAN, C. A. (1998) "The semantics of hierarchy : explicit parent-child relationship in MeSH tree structures", *Structures and relations in knowledge organization : proceedings of the Fifth International ISKO Conference, Lille, 25-29 August 1998*. Edited by W. Mustafa Elhadi, J. Maniez, S. Pollitt. Würzburg : Ergon Verlag, 1998. (Advances in knowledge organization 6), 133-115.

BEARMAN, D. et al. (1999) "A common model to support interoperable metadata : progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI Communities", *D-lib Magazine*, 5 (1) 1999. http://www.dlib.org/dlib/january99/bearman//01bearman.html.

BEAULIEU, M. (1997) "Experiments on interfaces to support query expansion", *Journal of Documentation*, 53 (1) 1997, 8-19

BEGHTOL, C. (1998) "General classification systems : structural principles for multidisciplinary specification", *Structures and relations in knowledge organization : proceedings of the Fifth International ISKO Conference, Lille, 25-29 August 1998*. Edited by W. Mustafa Elhadi, J. Maniez, S. Pollitt. Würzburg : Ergon Verlag, 1998. (Advances in knowledge organization 6), 88-95.

BELKIN, N. (1998) "An overview of results from Rutgers' investigations of interactive information retrieval", *Visualizing subject access for 21st century information resources : [proceedings of 34th Annual Clinic on Library Applications of Data Processing, Champaign, 1997]*. Edited by P. Atherton Cochrane, E. H. Johnson. Champaign, IL : Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 1998. 45-63.

BELL, B. L. (1998) An annotated guide to current national bibliographies. 2nd completely rev. ed. München : K. G. Saur, 1998. (UBCIM publications. New series 18).

BELL, B. L. (2001) "National bibliographies and the International Conference on National Bibliographic Services Recommendations : Africa, Middle East, and Asia", *67th IFLA Council and General Conference, Boston August 16-25, 2001*. http://www.ifla.org/IV/ifla67/papers/143-199e.pdf.

BENITO, M. (2001) "The UDC in Sweden", *Extensions & Corrections to the UDC*, 23 2001, 23-24.

BENITO, M. (2004) Udc translation survey. 2004-12-16. Message to mailing list udc-forum@jiscmail.ac.uk
http://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=ind04&L=udc-forum&T=0&F=&S=&P=2099.

BERNERS-LEE, T. (1996) "The World Wide Web : past, present and future" [Draft response to invitation to publish in IEEE Computer special issue of October 1996].
http://www.w3.org/People/Berners-Lee/1996/ppf.html.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. (2001) "The semantic web", *Scientific American*, 284 (5) 2001, 34-43.
http://www.sciam.com/2001/0501issue/0501berners-lee.html.

BHARAT, K.; BRODER, A. (1998) "A technique for measuring the relative size and overlap of public Web search engines", *Proceedings of the Seventh International Conference on World Wide Web 7, Brisbane, Australia, 1998*. Amsterdam : Elsevier, 1998, 379-388.
http://www7.conf.au/programme/fullpapers/1937/com1937.htm.

BHATTACHARYYA, G. (1979) "Fundamentals of subject indexing languages", *Ordering systems for global information networks : proceedings of the Third International Study Conference on Classification Research held at Bombay, India, during 6-11 January 1975*. Edited by A. Neelameghan. Bangalore : DRTC : FID/CR and Sarada Ranganathan Endowment for Library Science, 1979. (FID 533), 83-99.

BHATTACHARYYA, G. (1982) "Classaurus : its fundamentals, design and use", *Universal classification : subject analysis and ordering systems : proceeding of the 4th International Study Conference on Classification Research, 6th Annual Conference of Gesellshaft für Klassifikation, Augsburg, 28 June - 2 July 1982 : Vol. 1*. Edited by I. Dahlberg. Frankfurt : Indeks Verlag, 1982, 139-148.

BHATTACHARYYA, G.; RANGANATHAN, S. R. (1978) "From knowledge classification to library classification", *Conceptual basis of the classification of knowledge : proceedings of the Ottawa Conference, October 1st to 5th 1971*. Edited by J. A. Wojciechowski. New York; München; Paris : K.G. Saur, 1978. 119-143.

A Bibliographical Survey of UDC Editions (1982). The Hague : FID, 1982. (FID 573).

BINDING, C.; TUDHOPE, D. (2004) "KOS at your service : programmatic access to knowledge organization systems", *Journal of Digital Information*, 4 (4) 2004. http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Binding/.

BODOFF, D.; KAMBIL, A. (1998) "Partial coordination : I. The best of pre-coordination and post-coordination", *Journal of the American Society for Information Science*, 49 (14) 1998, 1254-1269.

BOKOS, G. D. (2000) "Bibliothikes stin Ellada : mia apopeira tekmiriomenis apotimisis tou topiou = Libraries in Greece : an attempt for a documented assessment of the field", *Tekmirion*, 2 2000, 133-230.

BORGMAN, C. (1996) "Why are online catalogs still hard to use?", *Journal of the American Society for Information Science*, 47 (7) 1996, 493-503.

BORGMAN, C. (1997) "From acting locally to thinking globally : a brief history of library automation", *Library Quarterly*, 67 1997, 215-219.

BORGMAN, C. et al. (1995) "Children's searching behavior on browsing and keyword online catalogs : The Science Library Catalog Project", *Journal of the American Society for Information Science*, 46 (9) 1995, 663-684.

BORKO, H. (1979) "Changing roles of and developments in automated classification", *Ordering systems for global information networks : proceedings of the Third International Study Conference on Classification Research held at Bombay, India, during 6-11 January 1975*. Edited by A. Neelameghan. Bangalore : DRTC : FID/CR and Sarada Ranganathan Endowment for Library Science, 1979. (FID 533), 214-219.

BORKO, H. (1982) "The role of classification in online retrieval systems and automated libraries", *Universal classification : subject analysis and ordering systems : proceeding of the 4th International Study Conference on Classification Research, 6th Annual Conference of Gesellshaft für Klassifikation, Augsburg, 28 June - 2 July 1982, Vol. 1*. Edited by I. Dahlberg. Frankfurt : Indeks Verlag, 1982, 235-245.

BOWKER, G. C. (1998) "The kindness of stranger : kinds and politics in classification systems", *Library Trends*, 47 (2) 1998, 254-292.

BOWKER, G. C.; STAR, S. L. (1999) Sorting things out : classification and its consequences. Cambridge, MA; London : The MIT Press, 1999.

BOWKER, L. (1998) "Peering through the linguistic keyhole : what can term choice tell us about knowledge organization?", *Structures and relations in knowledge organization : proceedings of the Fifth International ISKO Conference, Lille, 25-29 August 1998*. Edited by W. Mustafa Elhadi, J. Maniez, S. Pollitt. Würzburg : Ergon Verlag, 1998. (Advances in knowledge organization 6), 213-221.

BRADLEY, P. (1999) "The great search engine con-trick", *23rd International Online Information Meeting, London, 7-9 December 1999 : proceedings*. Edited by B. McKenna. Oxford : Learned Information Europe Ltd., 1999, 259-262.

BROUGHTON, V. (2000) "Structural, linguistic and mathematical elements in indexing languages and search engines : implications for the use of index language in electronic and non-LIS environments", *Dynamism and stability of knowledge organization : proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada*. Edited by C. Beghtol, L. C. Howarth, N. J. Williamson. Würzburg : Ergon Verlag, 2000. (Advances in knowledge organization 7), 206-212.

BROUGHTON, V.; LANE, H. (2000) "Classification schemes revisited : application to web indexing and searching", *Internet searching and indexing : the subject approach*. Edited by A. R. Thomas , J. R. Shearer, New York etc. : The Haworth Information Press, 2000, 143-155.

BROWNE, G. (1996) "Automatic indexing and abstracting", *Online Currents, the AusSI Newsletter,* 20 (6) 1996, 4-9*; LASIE,* 27(3) 1996, 58-65. http://www.aussi.org/conferences/papers/browneg.htm.

BUCKLAND, M. (1996) "The "liberal arts" of library and information science and the research university environment", *Proceedings CoLIS2 , Second International Conference on Conception of Library and Information Science : integration in perspective, October 13-16, 1996* . Edited by Peter Ingwersen, Niels Ole Pors. Copenhagen : The Royal School of Librarianship, 1996, 75-84.

BUCKLAND, M. (1999) "Mapping entry vocabulary to unfamiliar metadata vocabularies", *D-lib Magazine*, 5, (1) 1999. http://www.dlib.org/dlib/january99/buckland/01buckland.html.

BUCKLAND, M. et al. (1995) "Partnership in navigation : an information retrieval research agenda", *ASIS Annual Meeting, Chicago, October 1995*. http://www.asis.org/asis-95/papers/norgard.html.

BURNARD, L.; LIGHT, R. (1996) "Three SGML metadata formats : TEI, EAD, and CIMI", Study for BIBLINK, Work Package 1.1. December 1996 [project documentation] http://hosted.ukoln.ac.uk/biblink/wp1/sgml/.

BURTON, P. F. (1995) "Regulation and control of the Internet : is it feasible? Is it necessary?", *Journal of Information Science*, 21 (6), 1995, 413-428.

BUXTON, A. B. (1990) "Computer searching of UDC numbers", *Journal of Documentation*, 46 (3) 1990, 193-217. [An updated version published in: E*cyclopedia of library and information science*. Edited by A. Kent, H. Lancour. New York : Marcel Dekker, 1993, 51, 132-151.]

BUXTON, A. B.; HOPKINSON, A. (1994) The CDS/ISIS handbook. London : Library Association, 1994.

CALESS, T. W. (1969) "Subject analysis matrices for classification with UDC", *Proceedings of the First Seminar on UDC in a Mechanized Retrieval System conducted by R. R. Freeman and P. Atherton, Copenhagen, 2th[nd]-6th September, 1968*. Copenhagen : Danish Centre for Documentation, 1969. (FID/CR Report no 9), 29-37.

CALHOUN, K. (1998) "A bird's eye view of authority control in cataloging", *Proceedings of the Taxonomic Authority Files Workshop, Washington, DC, June 22-23, 1998*. http://research.calacademy.org/taf/proceedings/Calhoun.html.

CAMPBELL, G. (2000) "The relevance of traditional classification principles to the development and use of semantic markup languages for electronic text", *Dynamism and stability of knowledge organization : proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada*. Edited by C. Beghtol, L. C. Howarth, N. J. Williamson. Würzburg : Ergon Verlag, 2000. (Advances in knowledge organization 7), 345-351.

CAMPOCHIARO, E.; MARCELLI, F. (1991) "Thesaurus and UDC classification : an integrated system of on-line indexing and retrieval of parliamentary information", *Documentary languages and databases : papers from the Rome Conference, Dec 3-4, 1990*. Edited by G. Negrini, T. Farnesi, D. Benediktsson. Frankfurt/Main : Indeks Verlag, 1991. (Advances in knowledge organization 3), 161-167.

CAPLAN, P. (1995) "You call it corn, we call it syntax-independent metadata for document-like objects", *The Public Access Computer System Review,* 6 (5) 1995. http://info.lib.uh.edu/pr/v6/n4/capl6n4.html.

CAPLAN, P. (2003) Metadata fundamentals for all librarians. Chicago, IL : American Library Association, 2003.

CARLYLE, A. (1997) "The role of classification in the creation of author and work display in online catalogues", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 90-96.

Cataloging Internet resources (1997) : a manual and practical guide. 2nd ed. Edited by N. B. Olson. Dublin, OH : OCLC, 1997.

ČERNYJ, A. I. (1976) "O logičkim principima ustrojstva UDK i deskriptorskih jezika za pronalaženje informacije [The logical principles of the development of UDC and the information retrieval languages of descriptor type]", *Proceedings of the international symposium "UDC in Relation to Other Indexing Languages" held in Herceg Novi, Yugoslavia, June 28-July 1, 1971.* Beograd : Jugoslovenski centar za tehnicku i naucnu dokumentaciju; The Hague : FID, 1976, 23-36.

CHAN, L. M. (2000) "Exploiting LCSH, LCC, and DDC to retrieve networked resources : issues and challenges", Library of Congress, 2000. http://www.loc.gov/catdir/bibcontrol/chan_paper.html.

CHAN, L. M.; HODGES, T. L. (2000) "Library of Congress Classification", *The future of classification.* Edited by R. Marcella, A. Maltby. Aldershot : Gower, 2000, 105-127.

CHAN, L. M.; ZENG, M. L. (2004) "Ensuring interoperability among subject vocabularies and knowledge organization schemes", *Signum*, 8 2004, 25-30.

CHEN, H.; DHAR, V. (1990) "User misconceptions of information retrieval systems", *International Journal of Man-Machine Studies*, 32 (6) 1990, 673-692.

CHOWDHURY, G. G. (1999) "The Internet and information retrieval research : a brief review", *Journal of Documentation*, 55 (2) 1999, 209-225.

CHOWDHURY, G. G. (2003) Introduction to modern information retrieval. 2nd ed. London : Facet Publishing, 2003.

CHRISTENSEN, F. S. (2000) "Power and production of truth in the sciences", *Dynamism and stability of knowledge organization : proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada*. Edited by C. Beghtol, L. C. Howarth, N. J. Williamson. Würzburg : Ergon Verlag, 2000. (Advances in knowledge organization 7), 306-312.

CHU, H. (2003) Information representation and retrieval in the digital age. Medford, NJ : Information Today, 2003.

CLEVERDON, C. W.; MILLS, J. (1963) "The testing of index language devices", *Aslib Proceedings*, 15 (4) 1963, 106-130. [Reprinted in: From classification to "knowledge organisation, Dorking revisited or "past is prelude". Edited by A. Gilchrist. The Hague : FID, 1997. (FID 714), 65-83.]

COATES, E. J. (1970) "Switching language for indexing", *Journal of Documentation*, 26 (2) 1970, 102-11.

COATES, E. J.; LLOYD, G. A. ; SIMANDL, D. (1978) BSO - Broad System of Ordering : schedule and index. Prepared by the FID/BSO Panel, E. Coates, G. Lloyd, D. Simandl. The Hague : FID, 1978. (FID 564).

COATES, E. J.; LLOYD, G. A.; SIMANDL, D. (1979) The BSO manual : the development, rationale and use of the Broad System of Ordering. Prepared by the FID/BSO Panel, E. Coates, G. Lloyd, D. Simandl. The Hague : FID, 1979. (FID 580).

COCHRANE, P. A.; JOHNSON, E. J. (1996) "Visual Dewey : DDC in a hypertextual browser for the library user", *Knowledge organization and change :*

proceedings of the Fourth International ISKO Conference, 15-18 July 1996, Washington DC. Edited by R. Green. Frankfurt/Main : Indeks Verlag, 1996. (Advances in knowledge organization 5), 95-106.

COCHRANE, P. A.; MARKEY, K. (1985) "Preparing for the use of classification in online cataloging systems and in online catalogs", *Information Technology and Libraries*, 4 (2) 1985, 91-111.

Conceptual basis of the classification of knowledge : proceedings of the Ottawa Conference, October 1st to 5th 1971 = Les fondaments de la classification des savoirs : actes du Colloque d'Ottawa, du 1er au 5 octobre 1971. Edited by J. A. Wojciechowski. New York; München; Paris : K.G. Saur, 1978.

Concise UNIMARC Format for Classification Data. (2000) IFLA, PUC, 31 October, 2000. http://www.ifla.org/VI/3/p1996-1/concise.htm.

CORC builds on knowledge gained from InterCat and NetFirst. *OCLC Newsletter,* September/October, 235 1998.
http://www.oclc.org/oclc/new/n235/corc_builds_on_knowledge.htm.

CORDEIRO, M. I. (2003) "From library authority control to network authoritative metadata sources", *Subject retrieval in a networked environment : proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001*. Edited by I. C. McIlwaine. München : K. G. Saur, 2003. (UBCIM Publications - New Series 25), 131-139.

CORDEIRO, M. I.; SLAVIC, A. (2002) "Data models for knowledge organization tools : evolution and perspectives", *Challenges in knowledge representation and organization for the 21st century : integration of knowledge across boundaries : proceedings of the Seventh International ISKO Conference, 10-13 July 2002, Granada, Spain*. Edited by María J. López-Huertas with the assistance of Francisco J. Muñoz-Fernández. Würzburg : Ergon Verlag, 2002. (Advances in knowledge organization 8), 127-134.

COUSINS, S. A. (1992) "Enhancing subject access to OPACs : controlled vocabulary vs. natural language", *Journal of Documentation*, 48 (3) 1992, 291-309.

COVE, J. F.; WALSH, B.C. (1988) "Online text retrieval via browsing", *Information Processing and Management*, 24 (1) 1988, 31-37.

COVER, R. (2000) "Resource description and classification", *The XML Cover Pages*, 2000. http://www.oasis-open.org/cover/classification.html.

COX, K. (1992) "Information retrieval using browsing", [paper presented at the] 5th International Conference on New Information Technology, Nov. 30-Dec. 2, 1992 Hong Kong. http://web.simmons.edu/~chen/nit/NIT'92/069-cox.htm.

CRANEFIELD, S. (2001) "Networked knowledge representation and exchange using UML and RDF", *Journal of Digital Information*, 1 (8) 2001. http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Cranefield/.

CROFT, B. W. (1989) "Automatic indexing", *Indexing : the state of our knowledge and the state of our ignorance : proceedings of the 20th Annual Meeting of the American Society of Indexers*. Edited by B. H. Weinberg. Medford, NJ : Learned Information, 1989, 86-100.

CROFT, B. W. (1995) "What do people want from information retrieval", *D-Lib Magazine*, November 1995. http://www.dlib.org/dlib/november95/11croft.html.
[Reprinted in: From classification to "knowledge organisation" : Dorking revisited or "past is prelude". Edited by A. Gilchrist. The Hague : FID, 1997. (FID 714), 181-185.]

CROMWELL-KESSLER, W. (1998) "Crosswalks, metadata mapping and interoperability : what does it all mean?", *Introduction to metadata : pathways to digital information*. Edited by M. Baca. New York : Getty Information Institute, 1998, 19-22.

CROSS, P.; BRICKLEY, D.; KOCH, T. (2000) "Enhancement to user interface :

project deliverable", DESIRE project report, Work package 3, 2000. http://www.desire.org/html/research/deliverables/D3.6/d36b.html.

CROSS, P.; POWELL, A. (2000) "Guidelines on overcoming the problems of user interface design for large and/or distributed collection access. Guidelines for gateways on strategies to cope with the maintenance and evolution of large collections of resources : project deliverable ", *DESIRE project report, Work package 3, 2000*. http://www.desire.org/html/research/deliverables/D3.3/n33b.html.

CURRIER, S.; WAKE, S. (2001) "Negotiating subject access : resource discovery on the Web", *Library & Information Briefings*, 97 2001, 2-14.

DAHLBERG, I. (1971) "Possibilities for a new Universal Decimal Classification", *Journal of Documentation*, 27 (1) 1971, 18-36.

DAHLBERG, I. (1971a) "Possibilities for a reorganization of the UDC", *Proceedings of the Second Seminar on UDC in Mechanized Information Systems conducted by R. R. Freeman, Frankfurt, 1st -5th June, 1970.* Copenhagen : Danish Centre for Documentation, 1971. (FID/CR Report no 11), 193-210.

DAHLBERG, I. (1976) "UDK i idealni jezik za indeksiranje [The UDC and an ideal indexing language]", *Proceedings of the international symposium "UDC in Relation to Other Indexing Languages" held in Herceg Novi, Yugoslavia, June 28-July 1, 1971*. Beograd : Jugoslovenski centar za tehnicku i naucnu dokumentaciju; The Hague : FID, 1976, 37-55.

DAHLBERG, I. (1978) "Principles for the construction of a universal classification system : a proposal", *Conceptual basis of the classification of knowledge : proceedings of the Ottawa Conference, October 1st to 5th 1971*. Edited by J. A. Wojciechowski. New York; München; Paris : K.G. Saur, 1978. 450-457.

DAHLBERG, I. (1979) "On the theory of concepts", *Ordering systems for global information networks : proceedings of the Third International Study Conference on Classification Research held at Bombay, India, during 6-11 January 1975*. Edited by

A. Neelameghan. Bangalore : DRTC; FID/CR; Sarada Ranganathan Endowment for Library Science, 1979. (FID 533), 54-63.

DAHLBERG, I. (1991) "Philosophical foundations of conceptual ordering systems", *Documentary languages and databases : papers from the Rome Conference, Dec 3-4, 1990*. Edited by G. Negrini, T. Farnesi, D. Benediktsson. Frankfurt/Main : Indeks Verlag, 1991. (Advances in knowledge organization 3), 102-119.

DAHLBERG, I. (1992) "The basis of a new universal classification system seen from a philosophy of science point of view", *Classification research for knowledge representation and organization : proceedings of the 5th International Study Conference on Classification Research, Toronto, Canada, 24-28 June 1991*. Edited by N. J. Williamson, M. Hudon, Amsterdam : Elsevier Science Publishers; The Hague : FID, 1992, 187-209.

DAHLBERG, I. (1996) "Library catalogs on the Internet : switching for future subject access", *Knowledge organization and change : proceedings of the Fourth International ISKO Conference, 15-18 July 1996, Washington DC*. Edited by R. Green. Frankfurt/Main : Indeks Verlag, 1996. (Advances in knowledge organization 5), 155-164.

DAHLBERG, I. (1998) "Classification structure principles : investigations, experiences, conclusions", *Structures and relations in knowledge organization : proceedings of the Fifth International ISKO Conference, Lille, 25-29 August 1998*. Edited by W. Mustafa Elhadi, J. Maniez, S. Pollitt. Würzburg : Ergon Verlag, 1998. (Advances in knowledge organization 6), 79-82.

DAVIES, Ron (1996) "Thesaurus-aided searching in search and retrieval protocols", *Knowledge organization and change : proceedings of the Fourth International ISKO Conference, 15-18 July 1996, Washington DC.* Edited by R. Green. Frankfurt/Main : Indeks Verlag, 1996. (Advances in knowledge organization 5), 137-143.

DAVIES, Ron (2004) "Library portals - today and tomorrow", *Online information 2004: proceedings*.

http://www.online-information.co.uk/2004proceedings/thurspm/davies_r.pdf.

DAVIES, Roy (1989) "The creation of new knowledge by information retrieval and classification", *Journal of Documentation*, 45 (4) 1989, 273-301.

DAY, M. (1997) "Mapping Dublin Core to UNIMARC", UKOLN, Interoperability between metadata formats.
http://www.ukoln.ac.uk/metadata/interoperability/dc_unimarc.html.

DE GROLIER, E. (1961) "Points des vue retrospectif et prospectif dans la classification", *The Sayers memorial volume : essays in librarianship in memory of William Charles Berwick Sayers*. Edited by D. J. Foskett, B. I. Palmer for the Classification Research Group. London : The Library Association, 1961, 96-110.

DE GROLIER, E. (1991) "Some notes on the question of so-called 'unified classification'", *Tools for knowledge organization and the human interface : proceedings of the 1st International ISKO Conference, Darmstadt, 14-17 August 1990*. Edited by R. Fugmann. Frankfurt/Main : Indeks Verlag, 1991. (Advances in knowledge organization 2), 85-103.

DE GROLIER, E. (1991a) "Conference summary", *Tools for knowledge organization and the human interface : proceedings of the 1st International ISKO Conference, Darmstadt, 14-17 August 1990.* Edited by R. Fugmann. Frankfurt/Main : Indeks Verlag, 1991. (Advances in knowledge organization 2), 248-251.

DEKKERS, M; WEIBEL, S. (2002) "Dublin Core Metadata Initiative : progress report and workplan for 2002", *D-lib Magazine*, 8 (2) 2002.
http://www.dlib.org/dlib/february02/weibel/02weibel.html.

DEMPSEY, L. (1994) "Network resource discovery : a European library perspective", *Libraries, networks and Europe : a European networking study*. Edited by Neil Smith. London : British Library Research & Development, 1994. (Library and Information Research Paper 101), 29-68.
http://www.lub.lu.se/UB2proj/LIS_collection/lorcan.html.

DEMPSEY, L. (1996) "From ROADS to Desire", *D-lib Magazine*, July/August 1996. http://www.dlib.org/dlib/july96/07dempsey.html.

DEMPSEY, L. (2000) "The subject gateway : experiences and issues based on the emergence of the Resource Discovery Network", *Online Information Review*, 24 (1) 2000, 8-23.

DEMPSEY, L.; HEERY, R. (1998) "Metadata : a current view of practice and issues", *Journal of Documentation*, 54 (2) 1998, 145-172.

DEMPSEY, L.; WEIBEL, S. (1996) "The Warwick Metadata Workshop : a framework for the Deployment of Resource Description", *D-lib Magazine*, July/August 1996. http://www.dlib.org/dlib/july96/07weibel.html.

DENENBERG, R. (1996) "Structuring and indexing the Internet : keynote address", *Workshop on Earth Observation Catalogue Interoperability, 14-15 November 1996, Ispra, Varese, Italy*. http://www.loc.gov/z3950/agency/papers/italy.html.

DESIRE Information gateway handbook. (2000). http://www.desire.org/handbook/.

DEVADASON, F. J. (2003) "Faceted indexing application for organizing accessing Internet resources", *Subject retrieval in a networked environment : proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001*. Edited by I. C. McIlwaine. München : K. G. Saur, 2003. (UBCIM Publications - New Series 25), 147-159.

DEVADASON, F. J.; KOTHANDA RAMANUJAM, M. (1982) "Computer-aided construction of an 'alphabetic classaurus", *Universal classification : subject analysis and ordering systems : proceeding of the 4th International Study Conference on Classification Research, 6th Annual Conference of Gesellshaft für Klassifikation, Augsburg, 28 June - 2 July 1982 : Vol. 1*. Edited by I. Dahlberg. Frankfurt : Indeks Verlag, 1982, 173-179.

DEVADASON, F. J. et al. (2002) "Faceted Indexing Based System for Organizing and Accessing Internet Resources", *Knowledge Organization*, 29 (2) 2002, 65-77.

DEXTRE CLARKE, S; GILCHRIST, A.; WILL, L. (2004) "Revision and extension of thesaurus standards", *Knowledge Organization and the Global Information Society : proceedings of the 8th ISKO Conference, London 13-16 July 2004*. Edited by I. C. McIlwaine. Würzburg : Ergon Verlag, 2004. (Advances in knowledge organization 9), 215-220.

DING, Y. (2001) "A review of ontologies with the semantic web in view", *Journal of Information Science*, 27 (6) 2001, 377-384.

DIONGUE-DIOP, M. (1992) "Les problèmes de classification et d'indexation dans les bibliothèques africaines", *International Cataloguing and Bibliographic Control*, 21 (3) 1992, 44-47.

DOERR, M. (2001) "Semantic Problems of Thesaurus Mapping", *Journal of Digital Information*, 1 (8) 2001. http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr/.

DOERR, M. (2003) "The CIDOC CRM - an ontological approach to semantic interoperability of metadata", *AI Magazine*, 24 (3) (2003). http://cidoc.ics.forth.gr/docs/crm-ontology_preprint.pdf.

DREESE, M. J. (1979) "Comparison of document retrieval using UDC - notation given by UDC - expert with catchwords taken from title by clerk, stored in computer memory", *Ordering systems for global information networks : proceedings of the Third International Study Conference on Classification Research held at Bombay, India, during 6-11 January 1975*. Edited by A. Neelameghan. Bangalore : DRTC : FID/CR and Sarada Ranganathan Endowment for Library Science, 1979. (FID 533), 220-223.

Dublin Core Metadata Element Set, version 1.1 : reference description. (1999) DCMI 1999. http://dublincore.org/documents/1999/07/02/dces/

Dublin Core Metadata Initiative. http://dublincore.org/.

Dublin Core Qualifiers. (2000) DCMI 2000. http://purl.org/documents/rec/dcmes-qualifiers-200007.

DUBUC, R. (1973) La Classification Décimale Universelle (C.D.U.) : manuel pratique d'utilisation. 3rd ed. Paris; Bruxelles, Montréal : Gauthier-Villars, 1973.

DUNSIRE, G. (2000) "The Internet as a tool for cataloguing and classification, a view from the UK", *Internet searching and indexing : the subject approach*. Edited by A. R. Thomas , J. R. Shearer. New York etc : The Haworth Information Press, 2000, 187-195.

ELLIS, D. (1996) Progress and problems in information retrieval. London : Library Association Publishing, 1996.

ELLIS, D.; FORD, N. (1998) "In search of the unknown user : indexing, hypertext and the World Wide Web", *Journal of Documentation*, 54 (1) 1998, 28-47.

ELLIS, D.; VASCONCELOS, A. (1999) "Ranganathan and the Net : using facet analysis to search and organise the World Wide Web", *Aslib Proceedings*, 51 (1) 1999, 3-10.

ELLIS, D.; VASCONCELOS, A. (2000) "The relevance of facet analysis for World Wide Web subject organization and searching", *Internet searching and indexing : the subject approach*. Edited by A. R. Thomas, J. R. Shearer. New York etc. : The Haworth Information Press, 2000, 97-114.

Encoded Archival Description : EAD tag library for version 1.0 : Technical Document No. 2 (1998). Society of American Archivists and Library of Congress. June 1998. http://www.loc.gov/ead/tglib1998/tlhome.html.

FALLIS, D. (2000) "Consistency rules for classification schemes (or how to organize your beanie babies)", *Dynamism and stability of knowledge organization : proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto,* *Canada*. Edited by C. Beghtol, L. C. Howarth, N. J. Williamson. Würzburg : Ergon Verlag, 2000. (Advances in knowledge organization 7), 339-344.

ESTANBOULI, N. (1980) "Libraries in Syria", *Encyclopedia of library and information science*, Vol 29. Edited by A. Kent, H. Lancour. New York [etc.] : Marcel Dekker, 1980, 394-453.

FARRADANE, J. (1955) "The psychology of classification", *Journal of Documentation*, 11 (4) 1955, 187-201.

FARRADANE, J. (1961) "Fundamental fallacies and new need in classification", *The Sayers memorial volume* : *essays in librarianship in memory of William Charles Berwick Sayers*. Edited by D. J. Foskett, B. I. Palmer for the Classification Research Group. London : The Library Association, 1961, 120-135.

FENSEL, D. et al. (2000) "OIL in a nutshell", *Knowledge acquisition, modeling, and management : proceedings of the European Knowledge Acquisition Conference (EKAW-2000)*. Edited by R. Dieng et al. Würzburg : Springer-Verlag, 2000.

FILAKOVIĆ, Lj. (1981) "Kritike UDK i prijedlozi I. Dahlberg i Adolf-Friedrich Schmidta za reformu UDK [Criticism and suggestions for the reform of the UDC by I. Dahlberg and Adolf-Friedrich Schmidt]", *Vjesnik Bibliotekara Hrvatske*, 23 1981, 17-26.

First Seminar on UDC in a Mechanized Retrieval System (1969) conducted by R. R. Freeman and P. Atherton, Copenhagen, 2th[nd]-6th September, 1968 : proceedings. Edited by R. Molgaard-Hansen, M. Rigby. Copenhagen : Danish Centre for Documentation, 1969. (FID/CR Report no 9).

FOSKETT, A. C. (1973) The Universal Decimal Classification : the history, present status and future prospects of a large general classification scheme. London : Clive Bingley, 1973.

FOSKETT, A. C. (1997) The subject approach to information. 5th ed. London :

Library Association Publishing, 1997.

FOSKETT, A. C. (2000) "The future of faceted classification", *The future of classification*. Edited by R. Marcella, A. Maltby, Aldershot : Gower, 2000, 69-80.

FRÂNCU, V. (1996) "Building a multilingual thesaurus based on UDC", *Knowledge organization and change : proceedings of the Fourth International ISKO Conference, 15-18 July 1996, Washington DC.* Edited by R. Green. Frankfurt/Main : Indeks Verlag, 1996. (Advances in knowledge organization 5), 144-151.

FRÂNCU, V. (1997) "Language barriers and bridges : a comparative study on three UDC editions", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997.* The Hague : FID, 1997. (FID 716), 72-77.

FRÂNCU, V. (2000) "Harmonizing a universal classification system with an interdisciplinary multilingual thesaurus : advantages and limitations", *Dynamism and stability of knowledge organization : proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada*. Edited by C. Beghtol, L. C. Howarth, N. J. Williamson. Würzburg : Ergon Verlag, 2000. (Advances in knowledge organization 7), 200-205.

FRÂNCU, V. (2001) "A universal classification system going through changes", *Advances in classification research : Volume 10 : proceedings of the 10th ASIS Annual Meeting, November 1-5, 1999, Washington, DC*. Edited by H. Albrechtsen, J.-E. Mai. Medford, NJ : Information Today (for the American Society for Information Science and Technology), 2001. (ASIST Monograph Series), 55-72.

FRÂNCU, V. (2002) "Language-independent structure and multilingual information access", *Challenges in knowledge representation and organization for the 21st century : integration of knowledge across boundaries : proceedings of the Seventh International ISKO Conference, 10-13 July 2002, Granada, Spain.* Edited by María J. López-Huertas with the assistance of Francisco J. Muñoz-Fernández. Würzburg : Ergon Verlag, 2002. (Advances in knowledge organization 8), 404-411.

FRÂNCU, V. (2003) "The impact of specificity on the retrieval power of a UDC-based multilingual thesaurus", *Cataloging & Classification Quarterly*, 37 (1/2) 2003, 49 - 64.

FREDERKIN, R. et al. (1997) "Translingual information access", 1997 AAAI Spring Symposium on Cross Language Text and Speech Retrieval. http://www.cs.cmu.edu/~ref/aaai-ref.html.

FREEMAN, R. R. (1964) "Computers and classification systems", *Journal of Documentation* , 20 (3) 1964, 137-143.

FREEMAN, R. R. (1969) "Evaluation of the retrieval of metallurgical document references using UDC as the index language in a computer based system", *Proceedings of the First Seminar on UDC in a Mechanized Retrieval System conducted by R. R. Freeman and P. Atherton, Copenhagen, 2th[nd]-6th September, 1968*. Copenhagen : Danish Centre for Documentation, 1969. (FID/CR Report no 9)

FREEMAN, R. R.; ATHERTON, P. (1968) "File organization and search strategy using UDC in mechanized reference retrieval systems", *Mechanized Information Storage, Retrieval and Dissemination : proceedings of the FID/IFIP Joint Conference, Rome, June 14-17, 1967*. Edited by K. Samuelson. Amsterdam : North-Holland Publishing, 1968, 122-159.

FREEMAN, R. R.; ATHERTON, P. (1969) "Final report of the research project for the evaluation of the UDC as the indexing language for a mechanized retrieval system", *Proceedings of the First Seminar on UDC in a Mechanized Retrieval System conducted by R. R. Freeman and P. Atherton, Copenhagen, 2th[nd]-6th September, 1968*. Copenhagen : Danish Centre for Documentation, 1969. (FID/CR Report no 9), [29-37].

FREYRE, E.; NAUDI, M. (2003) "MACS : subject access across languages and networks", *Subject retrieval in a networked environment : proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001*. Edited by I. C. McIlwaine. München : K. G. Saur, 2003. (UBCIM Publications - New Series 25), 3-10.

FUGMANN, R. (1983) "The complementarity of natural and indexing languages", *Subject analysis and ordering systems : proceedings : 4th International Study Conference on Classification Research, Augsburg, 28 June-2 July 1982.* Edited by J. M. Perreault and I. Dahlberg. Frankfurt/Main : Indeks Verlag, 1983, 86-101.

FUGMANN, R. (1990) "Unused opportunities in indexing and classification", *Tools for knowledge organization and the human interface : proceedings of the 1st International ISKO Conference, Darmstadt, 14-17 August 1990.* Edited by R. Fugmann. Frankfurt/Main : Indeks Verlag, 1990. (Advances in knowledge organization 1), 65-77.

FUGMANN, R. (1993) Subject analysis and indexing : theoretical foundation and practical advice. Frankfurt/Main : Indeks Verlag, 1993. (Textbook for knowledge organization 1).

Functional requirements for authority records : a conceptual model [draft 18 Dec. 2003). IFLA UBCIM Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), 2003. [document distributed for comments to the members of the IFLA standing committees and groups on behalf of FRANAR Working Group via e-mail on 7th January 2004].

Functional requirements for bibliographic records : final report. (1998) IFLA Study Group on the Functional Requirements for Bibliographic records (FRBR), approved by the Standing Committee of the IFLA Section on Cataloguing. München : K. G. Saur, 1998.

GARCIA MARCO, F. J.; ESTABAN NAVARRO, M. A. (1993) "On some contributions of the cognitive science and epistemology to a theory of classification", *Knowledge Organization,* 20 (3) 1993, 126-132.

GARCIA MARCO, F. J.; ESTABAN NAVARRO, M. A. (1996) "Hypertext and indexing languages : common challenges and perspectives", *Knowledge organization and change : proceedings of the Fourth International ISKO Conference, 15-18 July 1996, Washington DC.* Edited by R. Green. Frankfurt/Main : Indeks Verlag, 1996.

(Advances in knowledge organization 5), 87-94.

General classification systems in a changing world : UDC and the Dewey centenary : proceedings of the FID Classification Symposium held in Commemoration of the Dewey Centenary, Brussels, November 1976. The Hague : FID, 1978. (FID 561).

GILCHRIST, A. (1972) "Intermediate language for switching and control", *Aslib Proceedings*, 24 (7) 1972, 387-399.

GILCHRIST, A. (1992) "UDC : the 1990s and beyond", *Classification research for knowledge representation and organization : proceedings of the 5th International Study Conference on Classification Research, Toronto, Canada, 24-28 June 1991*. Edited by N. J. Williamson, M. Hudon. Amsterdam : Elsevier Science Publishers; The Hague : FID, 1992, 69-78.

GILCHRIST, A. (2002) "From Aristotle to the 'semantic web'", *LA Record*, 104 (1) 2002. http://www.la-hq.org.uk/directory/record/r200201/article2.html.

GILCHRIST, A. (2003) "Thesauri, taxonomies and ontologies - an etymological note", *Journal of Documentation*, 59 (1) 2003, 7-18.

GILL, T. (1998) "Metadata and the World Wide Web", *Introduction to metadata : pathways to digital information*. Edited by M. Baca. New York : Getty Information Institute, 1998, 9-18.

GILLILAND-SWETLAND, A. J. (1998) "Defining metadata", *Introduction to metadata : pathways to digital information*. Edited by M. Baca. New York : Getty Information Institute, 1998, 1-8.

GNOLI, C.; RIDI, R.; VISINTIN, G. (2004) "Di che parla questo catalogo? : un'indagine sugli accessi semantici negli OPAC italiani", *Biblioteche oggi*, 22 (8) 2004, 23-29.

GOBYTĖ, M. (2000) "Martyndas Mazvydas National Library of Lithuania continues classification of documents using the UDC", *Extensions & Corrections to the UDC*, 22 2000, 32-34.

GÖDERT, W. (1991) "The design of subject access elements in online catalogues : some problems", *Tools for knowledge organization and the human interface : proceedings of the 1st International ISKO Conference, Darmstadt, 14-17 August 1990*. Edited by R. Fugmann. Frankfurt/Main : Indeks Verlag, 1991. (Advances in knowledge organization 2), 151-158.

GÖDERT, W. (1991a) "Facet classification in online retrieval", *International Classification*, 18 (2) 1991, 98-105.

GÖDERT, W. (1995) "Systematisches Suchen und Orientierung in Datenbanken", *Zwischen Schreiben und Lesen : Perspektiven für Bibliotheken, Wissenschaft und Kultur : Festschrift zum 60. Geburtstag von Hermann Havekost*. Edited by Hans-Joachim Wätjen. Oldenburg : Bibliotheks- und Informationssystem der Universität Oldenburg, 1995.
http://www.bis.uni-oldenburg.de/bisverlag/hv1/32-goede.pdf.

GÖDERT, W. (1996) "Vom OPAC zum Hyperkatalog : Suchen und Navigieren", *Erschließen, Suchen, Finden : Vorträge aus den bibliothekarischen Arbeitsgruppen der 19. und 20. Jahrestagungen (Basel 1995 / Freiburg 1996) der Gesellschaft für Klassifikation*. Edited by H.-J. Hermes u. H.-J. Wätjen. Oldenburg : Bibliotheks- und Informationssystem der Universität, 1996. 75-89.
http://www.bis.uni-oldenburg.de/bisverlag/herers96/kap2-3.pdf.

GÖDERT, W. (2003) "Navigation und Konzepte für ein interaktives Retrieval im OPAC oder Von der Informationserschließung zur Wissenserkundung", *AKMB-Veranstaltung Allegro und mehr... Bibliothekskataloge : Gestaltung und Mehrwertdienste, Wolfenbütte 20-21 Nov, 2003*.
http://voeb.uibk.ac.at/vm/vm57_1_goedert.pdf.

GÖDERT, W.; HORNY, S. (1990) "The design of subject access elements in online public access catalogue", *International Classification,* 17 (2) 1990, 66-76.

GOPINATH, M. A. (1997) "Frame based systems of mathematics on the classification systems", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997.* The Hague : FID, 1997. (FID 716), 161-166.

GOPINATH, M. A.; PRASAD, A. R. D. (1994) "A knowledge representation model for analytico-synthetic classification", *Knowledge organization and quality management : proceedings of the Third International ISKO Conference, 20-24 June 1994, Copenhagen, Denmark.* Edited by H. Albrechtsen, S. Oernager. Frankfurt/Main : Indeks Verlag, 1994. (Advances in knowledge organization 4), 320-327.

GORMAN, M. (1993) "How cataloguing and classification should be taught : a manifesto", *AACR, DDC, MARC and friends : the role of CIG in bibliographic control*. Edited by J. Byford, K. V. Trickey, S. Woodhouse. London : The Library Association, 1993, 111-125.

GREEN, D. (1999) "The evolution of Web searching", *23rd International Online Information Meeting, London, 7-9 December 1999 : proceedings*. Edited by B. McKenna. Oxford : Learned Information Europe Ltd., 1999. 251-258.

GUARINO, N. (1997) "Understanding, building and using ontologies", *International Journal of Human-Computer Studies*, 46 1997, 293-310. http://www.ladseb.pd.cnr.it/infor/Ontology/.

GUARINO, N.; WELTY, C. (2002) "Evaluating ontological decisions with OntoClean", *Communications of the ACM*, 45 (2) 2002, 61-65.

GUENTHER, R. S. (1992) "The USMARC format for classification data : development and implementation", *Classification research for knowledge representation and organization : proceedings of the 5th International Study Conference on Classification Research, Toronto, Canada, 24-28 June 1991*. Edited by N. J. Williamson, M. Hudon. Amsterdam : Elsevier Science Publishers; The Hague : FID, 1992, 235-245.

GUENTHER, R. S. (1994) "The Library of Congress Classification in the USMARC format", *Knowledge Organization,* 21 (4) 1994, 199-202.

GUENTHER, R. S. (1996) "Bringing the Library of Congress Classification into the computer age : converting LCC into machine-readable form", *Knowledge organization and change : proceedings of the Fourth International ISKO Conference, 15-18 July 1996, Washington DC*. Edited by R. Green. Frankfurt/Main : Indeks Verlag, 1996. (Advances in knowledge organization 5), 26-32.

Guidance information for the use of Dublin Core in Europe : draft [workshop documentation]. (2000) CEN/ISSS Workshop Metadata for Multimedia Information – Dublin Core, September 2000.

Guide to the Universal Decimal Classification (UDC) (1963) : B. S. 1000C : 1963. London : British Standards Institution, 1963.

Guidelines for the creation of a standard version of UDC (1991) : [FID internal documentation]. FID, UDC-CONS 91-54, UDC Ref. 91-5. March 1991.
[unpublished]

HAARALA, A. R. (1990) "FID pre-conference seminar on computer applications of the UDC", *The UDC : essays for a new decade*. Edited by A. Gilchrist, D. Strachan. London : Aslib, 1990, 33-34.

HAARALA, A. R. (1991) "The role of UDC in Finnish classification policy", *International Cataloguing and Bibliographic Control,* 20 (3) 1991, 43-46.

HAJDU BARÁT, A. (2004) "Knowledge organization of the Universal Decimal Classification", *Knowledge Organization and the Global Information Society : proceedings of the 8th ISKO Conference, London 13-16 July 2004.* Edited by I. C. McIlwaine. Würzburg : Ergon Verlag, 2004. (Advances in knowledge organization 9), 173-178.

HAKALA, J. (1998) "Dublin Core/FINMARC/GILS crosswalk : version 1.3".

Automation Unit of Finnish Research Libraries, Helsinki University Library, 1998. http://www.lib.helsinki.fi/meta/dcficross.html.

HALL, A. V. (1994) "Understanding the powerful grouping sense through a unifying classification theory", *Knowledge Organization,* 21 (3) 1994, 126-131.

HANCOCK, M. (1987) "Subject search behaviour at the library catalogue and at the shelves : implications for online interactive catalogues", *Journal of Documentation*, 43 (4) 1987, 303-321.

The handbook of research synthesis. (1994) Edited by H. Cooper, L. V. Hedges. New York : Russell Sage Foundation, 1994.

HARVEY, R. (1999) Organising knowledge in Australia : principles and practice in libraries and information centres. Wagga Wagga, N.S.W. : Centre for Information Studies, Charles Sturt University, 1999.

HENRIKSEN, T. (1998) "Description of indexing languages : a structural model". Oslo College, 1998. http://home.hio.no/~tor/descript.htm.

HEWSON, C. et al. (2003) Internet research methods : a practical guide for the social and behavioural sciences. London etc. : Sage Publications, 2003.

HICKEY, T. B.; VIZINE-GOETZ, D. (1999) "The role of classification in CORC", *23rd International Online Information Meeting, London, 7-9 December 1999 : proceedings*. Edited by B. McKenna. Oxford : Learned Information Europe Ltd., 1999, 247-250.

HIGH, W. M. (1990) "Library of Congress Classification numbers as subject access points in computer-based retrieval", *Cataloging and Classification Quarterly*, 11 (1) 1990, 37-43.

HILDRETH, C. R. (1991) "End users and structured searching of online catalogues : recent research finding", *Tools for knowledge organization and the human interface :*

proceedings 1st International ISKO Conference, Darmstadt, 14-17 August, 1990 : vol. 2. Edited by R. Fugmann. Frankfurt/Main : Indeks Verlag, 1991. (Advances in knowledge organization 2), 9-19.

HILDRETH, C. R. (1995) "The GUI OPAC : approach with caution" *The Public-Access Computer Systems Review*, 6 (5) 1995, 1-6.

HILDRETH, C. R. (1995a) Online catalog design models : are we moving in the right direction? : a report commissioned by The Council on Library Resources. August 1995. http://www.ou.edu/faculty/H/Charles.R.Hildreth/clr-five.html.

HILDRETH, C. R. (2000) "Are web-based OPACs more effective retrieval systems than their conventional predecessors? : an experimental study", *Dynamism and stability of knowledge organization : proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada*. Edited by C. Beghtol, L. C. Howarth, N. J. Williamson. Würzburg : Ergon Verlag, 2000. (Advances in knowledge organization 7), 237-246.

HILL, L.; KOCH, T. (2004) "Networked Knowledge Organization Systems : introduction to a special issue", *Journal of Digital Information*, 4 (4) 2004, article no. 53. http://jodi.ecs.soton.ac.uk/Articles/v01/i08/editorial/.

HILLMANN, D. (2003) "Using Dublin Core". DCMI Recommended resource, 2003. http://dublincore.org/documents/2003/08/26/usageguide/.

HIMANKA, J.; KAUTTO, V. (1993) "Translation of the Finnish abridged edition of UDC into general Finnish subject headings", *International Classification*, 19 (3) 1993, 131-134.

HINDSON, R. (1976) "Indeksiranje i pronalaženje podataka o britanskim specifikacijama za patente pomoću UDC i BISRA eksperimentalnog tezaurusa [Indexing and retrieval of records of British patent specifications using the Universal Decimal Classification and the BISRA experimental thesaurus]", *Proceedings of the international symposium "UDC in Relation to Other Indexing Languages" held in*

Herceg Novi, Yugoslavia, June 28-July 1, 1971. Beograd *:* Jugoslovenski centar za tehnicku i naucnu dokumentaciju; The Hague : FID, 1976, 61-110.

HINDSON, R. (1981) "UDC in the UK : a report on the 1979/1980 survey", *Aslib Proceedings*, 33 (3) 1981, 93-101.

HIOM, D. (1998) "Mapping classification schemes", SOSIG, ILRT, University of Bristol, 1998. http://www.sosig.ac.uk/desire/class/mapping.html.

HJERPPE, R. (1990) "The role of classification in hypertext : issues in implementing Roget's thesaurus in hypertext", *Tools for knowledge organization and the human interface : proceedings of the 1st International ISKO Conference, Darmstadt, 14-17 August 1990*. Edited by R. Fugmann. Frankfurt/Main : Indeks Verlag, 1990. (Advances in knowledge organization 1), 206-215.

HJERPPE, R. (1996) "Go with a flow, or abide by the side, or watch the waves? : challenges of change for knowledge organization", *Knowledge organization and change : proceedings of the Fourth International ISKO Conference, 15-18 July 1996, Washington DC.* Edited by R. Green. Frankfurt/Main : Indeks Verlag, 1996. (Advances in knowledge organization 5), 10-25.

HJØRLAND, B. (1994) "Nine principles of knowledge organization", *Knowledge organization and quality management : proceedings of the Third International ISKO Conference, 20-24 June 1994, Copenhagen, Denmark.* Edited by H. Albrechtsen, S. Oernager. Frankfurt/Main : Indeks Verlag, 1994. (Advances in knowledge organization 4), 91-100.

HJØRLAND, B. (1997) Information seeking and subject representation : an activity-theoretical approach to information science. Westport, CT : Greenwood Press, 1997.

HJØRLAND, B. (2003) "Fundamentals of knowledge organization", *Knowledge Organization,* 30 (2) 2003, 87-111.

HODGE, G. (2000) Systems of knowledge organization for digital libraries : beyond traditional authority files. Council of Library and Information Resources, April 2000. http://www.clir.org/pubs/reports/pub91/contents.html.

HORTON, F. W. (2003) "International Federation for Information and Documentation", *Encyclopedia of library and information science*, Vol 2. 2nd ed. Edited by M. A. Drake. New York [etc.] : Marcel Dekker, 2003. 1405-1407.

HOWARD, L. (1990) "Call numbers and collating sequences", *LASIE*, 20 (5-6) 1990, 84-98.

HOWARTH, L. C. (2000) "Designing a 'human understandable' metalevel ontology for enhancing resource discovery in knowledge bases", *Dynamism and stability of knowledge organization : proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada*. Edited by C. Beghtol, L. C. Howarth, N. J. Williamson. Würzburg : Ergon Verlag, 2000. (Advances in knowledge organization 7), 391-397.

How to obtain the UDC? *Extensions & Corrections to the UDC*, 26 2004, 45.

HUG, H.; NÖTHIGER, R. (1988) "ETHICS : an online public access catalogue at ETH Bibliothek, Zurich", *Program*, 22 (2) 1988, 133-142.

HUG, H.; WALSNER, M. (1990) "Retrieval in the ETH database using the UDC", *Tools for knowledge organization and the human interface : proceedings of the 1st International ISKO Conference, Darmstadt, 14-17 August 1990*. Edited by R. Fugmann. Frankfurt/Main : Indeks Verlag, 1990. (Advances in knowledge organization 1), 216-219.

HUNTER, E. J. (2000) "Do we still need classification?", *The future of classification*. Edited by R. Marcella, A. Maltby. Aldershot : Gower, 2000, 1-31.

HUNTER, J. (2001) "MetaNet - A metadata term thesaurus to enable semantic interoperability between metadata", *Journal of Digital Information*, 1 (8) 2001. http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Hunter/.

HUTCHINS, W. J. (1975) Languages of indexing and classification. Stevenage : Peter Peregrinus, 1975.

HUXLEY, L. (2000) "New features of The Social Science Information Gateway", *Proceedings of the JANET User Support Workshop, 4th-6th July 2000*. Cambridge : University of Cambridge, 2000, 99-107.
http://www.sosig.ac.uk/about_us/docs/sosig_paper.pdf.

IEEE Information Technology - Learning Technology - Learning Object Metadata. (2000) Draft Standard for Learning Object Metadata : IEEE P1484.12/D4.0, 5 February 2000. http://ltsc.ieee.org/doc/wg12/LOM_WD4.htm.

IFIDON, S. E. (1974) "Special problems facing African librarians : the West African experience", *Libri*, 24 (4) 1974, 310-318.

IHADJADENE, M.; BOUCHÉ, R. (1998) "Using syntagmatic relationships based on a RAMEAU list as a browsing relevance feedback strategy in a WWW-OPAC", *Structures and relations in knowledge organization : proceedings of the Fifth International ISKO Conference, Lille, 25-29 August 1998*. Edited by W. Mustafa Elhadi, J. Maniez, S. Pollitt. Würzburg : Ergon Verlag, 1998. (Advances in knowledge organization 6), 161-167.

IHADJADENE, M.; BOUCHÉ, R.; ZAAFRANI, R. (2000) "The dynamic nature of searching and browsing on web OPACs : the CATHIE experience", *Dynamism and stability of knowledge organization : proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada*. Edited by C. Beghtol, L. C. Howarth, N. J. Williamson. Würzburg : Ergon Verlag, 2000. (Advances in knowledge organization 7), 327-331.

IMS Learning Resource Meta-data Information Model : version 1.1 - Final Specifications. IMS Global Learning Consortium, 2001.
http://www.imsproject.org/metadata/mdinfov1p1.html.

Information, knowledge, evolution : proceedings of the Forty-fourth FID Congress held in Helsinki, Finland, 28 August-1 September, 1988. Edited by S. Koskiala, R. Launo. Amsterdam : North-Holland, 1989. (FID 675).

Introduction to metadata (1998) : pathways to digital information. Edited by M. Baca. New York : Getty Information Institute, 1998.

ISHIKAWA, T. (1994) "UDC number automatic combination system (UDC-AUTCS)", *Knowledge organization and quality management : proceedings of the Third International ISKO Conference, 20-24 June 1994, Copenhagen, Denmark.* Edited by H. Albrechtsen, S. Oernager. Frankfurt/Main : Indeks Verlag, 1994. (Advances in knowledge organization 4), 328-333.

ISO 2788, Documentation - guidelines for the establishment and development of monolingual thesauri : international standard. (1974) Geneva : International Organization for Standardization, 1974.

ISO 5963, Documentation - methods for examining documents, determining their subjects, and selecting indexing terms : international standard. (1985) Geneva, International Organization for Standardization, 1985.

ISO/IEC 13250:2000 Topic Maps: Information Technology -- Document Description and Markup Languages (1999). Edited by M. Biezunski, M. Bryan, S. R. Newcomb. First, December 3 1999.
http://www.y12.doe.gov/sgml/sc34/document/0129.pdf.

JACOB, E. K. (1994) "Classification and crossdisciplinary communication : breaching boundaries imposed by classification structure", *Knowledge organization and quality management : proceedings of the Third International ISKO Conference, 20-24 June 1994, Copenhagen, Denmark.* Edited by H. Albrechtsen, S. Oernager. Frankfurt/Main : Indeks Verlag, 1994. (Advances in knowledge organization 4), 101-108.

JACOBSON, F. F.; IGNACIO, E. N. (1997) "Teaching reflection : information seeking and evaluation in the digital library environment", *Library Trends*, 46, 1997, 771-802.

JARVIS, H. (1993) "Restoring the bibliographic heritage of Vietnam and Cambodia", *International Cataloguing and Bibliographic Control*, 22 (3) 1993, 42-45.

JENKINS, C. et al. (1998) "Automatic classification of web resources using Java and Dewey Decimal Classification", *Computer Networks and ISDN Systems*, 30 1998, 646-648. http://www.scit.wlv.ac.uk/~ex1253/classifier/.

JUL, E. (1997) "Cataloging Internet resources : survey and prospectus", *Bulletin of the American Society for Information Science*, October 1997.
http://www.asis.org/Bulletin/Oct-97/jul.htm.

JURISICA, I. (2000) "Knowledge organization by systematic knowledge management and discovery", *Dynamism and stability of knowledge organization : proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada*. Edited by C. Beghtol, L. C. Howarth, N. J. Williamson. Würzburg : Ergon Verlag, 2000. (Advances in knowledge organization 7), 366-371.

KAJOSALO, E. (1997) "Issues related to cataloging of Internet resources", *A paper prepared for LIS598 - Application of Technology in Libraries at the University of Alberta*, March 20, 1997 http://hollyhock.slis.ualberta.ca/598/erja/rep_html.

KARA MAROSI, A. (1976) "Odnos izmedju UDK i tezaurusa EVROATOM-a sa gledišta male informacijske službe [The relation of the UDC to the EURATOM thesaurus from the point of view of the small information centre]", *Proceedings of the international symposium "UDC in Relation to Other Indexing Languages" held in Herceg Novi, Yugoslavia, June 28-July 1, 1971*. Beograd : Jugoslovenski centar za tehnicku i naucnu dokumentaciju; The Hague : FID, 1976, 111-122.

KARAMAĆ, B. (1991) "Bibliographic control in Poland", *International Cataloguing and Bibliographic Control*, 20 (1), 1991, 3-7.

KARHULA, P. (1990) "Use and usability of the UDC in classification practice and online retrieval", *The UDC : essays for a new decade*. Edited by A. Gilchrist and D. Strachan. London : Aslib, 1990, 47-53.

The KBS-Cross Project, Knowledge-Based System for automatic CROSS-referencing of classification systems - KBS-CROSS. (1995) Lund University, 1995. http://it.civil.auc.dk/it/delphi/KBS/projects/kbscross.html.

KELLEY, D. (1998) The art of reasoning : with symbolic logic. Expanded edition. NewYork : W.W. Norton and Co, 1998.

KEMP, D.A. (1976) The nature of knowledge : an introduction for librarians. London : Clive Bingley, 1976.

KENT, R. E. (1998) "Organizing conceptual knowledge online : metadata interoperability and faceted classification", *Structures and relations in knowledge organization : proceedings of the Fifth International ISKO Conference, Lille, 25-29 August 1998*. Edited by W. Mustafa Elhadi, J. Maniez, S. Pollitt. Würzburg : Ergon Verlag, 1998. (Advances in knowledge organization 6), 388-395.

KILGOUR, F. G. (1997) "Origins of coordinate searching", *Journal of the American Society for Information Science,* 48 (4) 1997, 340-348.

KIRRIEMUIR, J. (1999) "A brief survey of Quality Resource Discovery Systems : commissioned by the JISC-funded Resource Discovery Network Centre : final report". JISC RDN, September 1999.
http://www.rdn.ac.uk/publications/studies/survey/.

KIRRIEMUIR, J. et al. (1998) "Cross-searching subject gateways : the query routing and forward knowledge approach", *D-lib Magazine*, January 1998. http://www.dlib.org/january98/01kirriemuir.html.

KIRSEL, A.-M. (1994) "UDC and Estonia", *Extensions & Corrections to the UDC*, 16 1994, 31-32.

KNIESNER, D. L.; WILLMAN, C. (1995) "But is it an online shelflist? : classification access in eight OPACs", *Cataloging & Classification Quarterly*, 20 (4) 1995, 5-21.

KOCH, T. (2000) "Quality-controlled subject gateways : definitions, typologies, empirical overview", *Online Information Review*, 241, 2000.
http://www.lub.lu.se/~traugott/OIR-SBIG.txt.

KOCH, T. (2000a) "Dublin Core Metadata Initiative in transition : report", *DC 8, Ottawa, 3-6 October 2000*. http://www.lub.lu.se/metadata/dc8-report.html.

KOCH, T. (2003) "Controlled vocabularies, thesauri and classification systems available in the WWW : DC Subject". NetLab, Lund University Library Development Department, Sweden, 2003. http://www.lub.lu.se/metadata/subject-help.html.

KOCH, T. et al. (1997) "The role of classification schemes in Internet resource description and discovery : project final report". DESIRE project, 1997.
http://www.lub.lu.se/desire/radar/reports/D3.2.3/.

KOCH, T. et al. (2000) "Provide browsing using classification schemes", DESIRE Information gateway handbook, 2000.
http://www.lub.lu.se/desire/handbook/class.html.

KOCH, T.; ARDÖ, A.; GOLUB, K. (2004) "Browsing and searching behaviour in the Renardus Web Service : a study based on log analysis", [report presented at JCDL'04, June 7-11, 2004, Tucson, Arizona, USA].

KOCH, T; NEUROTH, H.; DAY, M. (2003) "Renardus : cross-browsing European subject gateways via a common classification system (DDC)", *Subject retrieval in a networked environment : proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001*. Edited by I. C. McIlwaine. München : K. G. Saur, 2003. (UBCIM Publications - New Series 25), 25-33.

KOCH, T.; VIZINE-GOETZ, D. (1998) "Automatic classification and content navigation support for web services : DESIRE II cooperates with OCLC". OCLC, Dublin, OH ; NetLab, Lund University Library Development Department, Sweden. 1998. http://digitalarchive.oclc.org/da/ViewObject.jsp?objid=0000003489.

KOH, G. S. (1995) "Options in classification available through modern technology", *Cataloging & Classification Quarterly*, 19 (3-4) 1995, 195-211.

KUMAR, T. V. R.; PARAMESWARAN, M. (1999) "Chain indexing and UDC ", *Annals of Library Science and Documentation*, 46 (2) 1999, 53-58.

KURHULA, P. (1990) "Use and usability of the UDC in classification practice and online retrieval", *The UDC : essays for a new decade*. Edited by A. Gilchrist, D. Strachan. London : Aslib, 1990, 48-55.

KWASNIK, B. H. (1999) "The role of classification in knowledge representation and discovery", *Library Trends*, 48 (1) 1999, 22-47.

KWASNIK, B. H.; LIU, X. (2000) "Classification structures in the changing environment of active commercial websites : the case of eBay.com", *Dynamism and stability of knowledge organization : proceedings of the Sixth International ISKO Conference*, *10-13 July 2000, Toronto, Canada*. Edited by C. Beghtol, L. C. Howarth, N. J. Williamson. Würzburg : Ergon Verlag, 2000. (Advances in knowledge organization 7), 372-377.

KYLE, B. (1961) "The UDC : a study of the present position and of its future developments, with particular attention to those schedules which deal with the humanities, arts, and social sciences". *UNESCO Bulletin for Libraries*, 15 (2) 1961, 53-69.

LA BARRE, K. (2004) "Adventures in faceted classification : a brave new world or a world of confusion", *Knowledge Organization and the Global Information Society : proceedings of the 8th ISKO Conference, London 13-16 July 2004*. Edited by I. C. McIlwaine. Würzburg : Ergon Verlag, 2004. (Advances in knowledge organization

9), 79-84.

LAFFORD, A.; STONE, O. (1997) "NISS - National Information Services and Systems", *Ariadne*, 10 1997. http://www.ariadne.ac.uk/issue10/niss/.

LAGOZE, C. (1996) "The Warwick Framework", *D-lib Magazine*, July/August 1996. http://www.dlib.org/dlib/july96/lagoze/07lagoze.html.

LAGOZE, C. (2000) "Accommodating simplicity and complexity in metadata : lessons from the Dublin Core experience", [presented at] The Seminar on Metadata at Netherlands Institute for Archival Education and Research, June 8 2000. http://www.cs.cornell.edu/lagoze/Papers/dc.pdf.

LAGOZE, C. (2001) "Keeping Dublin Core simple : cross-domain discovery or resource description?", *D-lib Magazine*, 7 (1) 2001. http://www.dlib.org/dlib/january01/lagoze/01lagoze.html.

LANCASTER, F. W.; SANDORE, B. (1997) Technology and management in library and information services. London : Library Association Publishing, 1997.

LANGRIDGE, D. W. (1976) Classification and indexing in the humanities. London : Butterworth, 1976.

LAROUK, O. (1998) "Modeling users needs : schemas of interrogation and filtering of answers from the web in co-operative mode", *Structures and relations in knowledge organization : proceedings of the Fifth International ISKO Conference, Lille, 25-29 August 1998*. Edited by W. Mustafa Elhadi, J. Maniez, S. Pollitt. Würzburg : Ergon Verlag, 1998. (Advances in knowledge organization 6), 106-107.

LASSILA, O. (1997) "Introduction to RDF Metadata : W3C NOTE 1997-11-13", W3C Technology and Society Domain : Resource Description Framework. http://www.w3.org/TR/NOTE-rdf-simple-intro-971113.html.

LASSILA, O.; MCGUINNESS, D. (2001) "The role of frame-based representation on the Semantic Web", Linköping Electronic Articles in Computer and Information Science, 6 (5) 2001. http://www.ida.liu.se/ext/epa/cis/2001/005/paper.pdf.

LAWRENCE, S; GILES, C. L. (1998) "Searching the World Wide Web", *Science*, 280, 1998, 98-100.

LEBLANC, J. (1995) "Classification and shelflisting as value added; some remarks on the relative worth and price of predictability, serendipity, and depth of access", *Library Resources and Technical Services*, 39 (3) 1995, 294-302.

LEE PAO, M. (1989) Concepts of information retrieval. Englewood, CO : Libraries Unlimited, 1989.

LESK, M. (1995) "The seven ages of information retrieval : UDT occasional paper 5", *IFLA Universal Dataflow and Telecommunications (UDT)* : *occasional papers, 1995*. http://www.ifla.org/VI/5/op/udtop5/udtop5.htm.

LEŠČIĆ, J; CVITAŠ, M. (2003) "UDC based subject-analysis system", *Extensions & Corrections to the UDC*, 25 2003, 41-42.

LEWIS, D. D.; SPARCK JONES, K. (1996) "Natural language processing for information retrieval", *Communications of the ACM (CACM)*, 39 (1) 1996, 92-101. [Reprinted in: From classification to "knowledge organisation" : Dorking revisited or "past is prelude". Edited by A. Gilchrist. The Hague : FID, 1997. (FID 714), 49-61.]

LIU, S. (1990) "Online classification notation: proposal for a flexible faceted notation system", *International Classification*, 17 (1) 1990, 14-20.

LIU, S. (1996) "Decomposing DDC synthesized numbers", *62nd IFLA General Conference Beijing, China, August 25-31, 1996*. http://ifla.org/IV/ifla62/62-sonl.htm.

LLOYD, G. A. (1969) "UDC in its international aspects", *Aslib Proceedings*, 21 (5) 1969, 204-208.

LLOYD, G. A. (1971) "Why this Seminar? : keynote address to the Seminar", *Proceedings of the Second Seminar on UDC in Mechanized Information Systems conducted by R. R. Freeman, Frankfurt, 1st -5th June, 1970*. Copenhagen : Danish Centre for Documentation, 1971. (FID/CR Report no 11),1- 2.

LLOYD, G. A. (1972) "UDC : revise or relegate? : FID's Standard Reference Code project and UDC improvement programme", *Aslib Proceedings*, 24 (10) 1972, 580-587.

LONG, C. E. (2000) "Improving subject searching in Web-based OPACs : evaluation of the problem and guidelines for design", *Internet searching and indexing : the subject approach*. Edited by A. R. Thomas , J. R. Shearer. New York etc : The Haworth Information Press, 2000, 159-186. [Also published in *Journal of Internet Cataloguing*, 2 (3/4) 2000, 159-186.]

LOSEE, R. M. (1993) "Seven fundamental questions for the science of library classification", *Knowledge Organization,* 20 (2) 1993, 65-70.

LOTH, K. (1996) "Überlegungen zu einer computergerechten Reorganisation der UDK", *Extensions & Corrections to the UDC*, 18 1996, 9-13.

LOTH, K.; FUNK, H. (1990) "Subject search on ETHICS on the basis of UDC", *The UDC : essays for a new decade*. Edited by A. Gilchrist, D. Strachan. London : Aslib, 1990, 35-47.

LYNCH, C. (1995) "Networked information resource discovery : an overview of current issues", *IEEE Journal on selected areas in communications*, 13 (8) 1995, 1505-1522.

McCASH, W. H.; CARMICHAEL, J. (1970) "UDC user profiles as developed for a computer-based SDI service in the Iron and Steel Industry", *Journal of Documentation*, 26 (4) 1970, 295-312.

MacDOUGALL, S. (2000) "Signpost of the information superhighway : indexes and access", *Internet searching and indexing : the subject approach*. Edited by A. R. Thomas , J. R. Shearer. New York etc : The Haworth Information Press, 2000, 61-79.

McGARRY, D. (1992) "Displays of bibliographic records in call number order : functions of the displays and data elements needed", *Cognitive paradigms in knowledge organization : [proceedings of the] Second International ISKO conference, Madras, 26-28 August 1992*. Bangalore : Sarada Ranganathan Endowment for Library Science, 1992. 65-94.

McGARRY, D. (1997) "Mathematics in four classification schemes : a discussion", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 78-84.

McGUINNESS, D. L. (2002) "Ontologies come of age", *Spinning the semantic web : bringing the World Wide Web to its full potential*. Edited by D. Fensel et al. Cambridge, MA : MIT Press, 2002, 171-194. http://www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-(with-citation).htm.

McILWAINE, I. C. (1990) "The work of the System Development Task Force", *The UDC : essays for a new decade*. Edited by A. Gilchrist, D. Strachan. London : Aslib, 1990. 19-28.

McILWAINE, I. C. (1993) "UDC : the present state and future development", *59th IFLA Council and Conference, Barcelona, Spain 22-28 August 1993*, booklet 4, 37-39.

McILWAINE, I. C. (1995) "UDC centenary : the present state and future prospects", *Knowledge Organization*, 22 (2) 1995, 64-69.

McILWAINE, I. C. (1996) "New wine in old bottles : problems of maintaining classification schemes", *Knowledge organization and change : proceedings of the Fourth International ISKO Conference, 15-18 July 1996, Washington DC*. Edited by

R. Green. Frankfurt/Main : Indeks Verlag, 1996. (Advances in knowledge organization 5), 122-136.

McILWAINE, I. C. (1997) "Classification and linear orders : problems of organizing Zoology", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 134-138.

McILWAINE, I. C. (1997a) "The Universal Decimal Classification : some factors concerning its origins, development, and influence", *Journal of the American Society for Information Science*, 48 (4) 1997, 331-339. [Published in a revised version in : *Historical studies in information science*. Edited by T. B. Hahn and M. Buckland. Medford, NJ : Information Today, 1998, 94-106]

McILWAINE, I. C. (1998) "Knowledge classification, bibliographic classification and the Internet", *Structures and relations in knowledge organization : proceedings of the Fifth International ISKO Conference, Lille, 25-29 August 1998*. Edited by W. Mustafa Elhadi, J. Maniez, S. Pollitt. Würzburg : Ergon Verlag, 1998. (Advances in knowledge organization 6), 96-104.

McILWAINE, I. C. (1998a) "The UDC at the present time", *64th IFLA [Council and] General Coference, Amsterdam 16-21 August 1998*.
http://www.ifla.org/IV/ifla64/102-158e.htm.

McILWAINE, I. C. (1998b) "UDC into the 21st century", *Aslib Proceedings*, 50 (2) 1998, 44-48.

McILWAINE, I. C. (2000) "Interdisciplinarity: a new retrieval problem?", *Dynamism and stability of knowledge organization: proceedings of the Sixth International ISKO Conference,10-13 July 2000, Toronto, Canada*. Edited by C. Beghtol, L. C. Howarth, N. J. Williamson. Würzburg: Ergon Verlag, 2000. (Advances in knowledge organization 7), 261-267.

McILWAINE, I. C. (2000a) "UDC in the twenty-first century", *The future of* *classification*. Edited by R. Marcella, A. Maltby. Aldershot : Gower, 2000, 93-104.

McILWAINE, I. C. (2000b) The Universal Decimal Classification : a guide to its use. 3rd ed. The Hague : UDC Consortium, 2000. (UDC Publication, no P035).

McILWAINE, I. C. (2001) "Prilagodba UDK mrežnom svijetu [Adapting the UDC to the network community]", *Vjesnik bibliotekara Hrvatske*, 44 (1-4) 2001, 52-67.

McILWAINE, I. C. (2003) "The UDC and the World Wide Web", *Subject retrieval in a networked environment : proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001*. Edited by I. C. McIlwaine. München : K. G. Saur, 2003. (UBCIM Publications - New Series 25), 170-176.

McILWAINE, I. C.; WILLIAMSON, N. J. (1994) "A feasibility study on the restructuring of the Universal Decimal Classification into a fully-faceted classification system", *Knowledge organization and quality management : proceedings of the Third International ISKO Conference, 20-24 June 1994, Copenhagen, Denmark*. Edited by H. Albrechtsen, S. Oernager. Frankfurt/Main : Indeks Verlag, 1994. (Advances in knowledge organization 4), 406-413.

McILWAINE, I. C.; WILLIAMSON, N. J. (1999) "International trends in subject analysis research", *Knowledge Organization*, 26 (1) 1999, 23-29.

McKIERNAN, G. (1997) "The big picture : visual browsing in Web and non-Web databases". 1997. http://www.public.iastate.edu/~CYBERSTACKS/DPC97.htm.

MacLENNAN, A (2000) "Classification and the Internet", *The future of classification*. Edited by R. Marcella, A. Maltby. Aldershot : Gower, 2000, 59-68.

MacMILLEN, S. (1992) "Vietnam, libraries and information services", *Encyclopedia of library and information science*. Edited by A. Kent, H. Lancour. New York : Marcel Dekker, 1992, 50, 363-389.

MACEVICIUTE, E. (1995)"International bibliographic standards in Lithuania : the impact of decentralization", *Library Management*, 16 (8) 1995, 42-46.

MADALLI, D. P.; PRASAD, A. R. D. (2002) "VYASA : a knowledge representation system for automatic maintenance of analytico-synthetic scheme", *Challenges in knowledge representation and organization for the 21st century : integration of knowledge across boundaries : proceedings of the Seventh International ISKO Conference, 10-13 July 2002, Granada, Spain*. Edited by María J. López-Huertas with the assistance of Francisco J. Muñoz-Fernández. Würzburg : Ergon Verlag, 2002. (Advances in knowledge organization 8), 113- 119.

MAI, J.-E. (1997) "The concept of subject : on problems in indexing", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 60-66.

MAI, J.-E. (2003) "The future of general classification", *Cataloging & Classification Quarterly*, 37 (1/2) 2003, 3-12.

MALTBY, A; MARCELLA, R. (2000) "Organizing knowledge : the need for system and unity", *The future of classification*. Edited by R. Marcella, A. Maltby. Aldershot : Gower, 2000, 19-31.

MANDEL, C. (1995) "Change and continuity in subject authority control", *The future is now : reconciling change and continuity control : proceedings of the OCLC Symposium, ALA Annual Conference, June 23, 1995*.
http://www.oclc.org/oclc/man/9391ausy/mandle.htm.

MANIEZ, J. (1997) "Database merging and the compatibility of indexing languages", *Knowledge Organization*, 24 (4) 1997, 213-223.

MANIEZ, J. (1991) "Are classifications still relevant in databases?", *Documentary languages and databases : papers from the Rome Conference, Dec 3-4, 1990*. Edited by G. Negrini, T. Farnesi, D. Benediktsson. Frankfurt/Main : Indeks Verlag, 1991. (Advances in knowledge organization 3), 120-129.

MARC 21 Concise Format for Classification Data. (2003) Concise edition. Library of Congress, 2003. http://www.loc.gov/marc/classification/eccdhome.html.

MARCELLA, R.; NEWTON, R. (1994) A new manual of classification. Aldershot; Brookfield, VT : Gower, 1994.

MARCHIONINI, G. (1995) Information seeking in electronic environment. Cambridge : Cambridge University Press, 1995.

MARCHIONINI, G.; BRUNK, B. (2003) "Towards a general relation browser : a GUI for information architects", *Journal of Digital Information*, 4 (1) 2003. http://jodi.ecs.soton.ac.uk/Articles/v04/i01/Marchionini/.

MARCO, D. (2000) Building and managing the meta data repository : a full lifecycle guide. New York, etc : John Wiley & Sons, 2000.

MARKER, R. J. (1996) "The role and function of the shelflist in an automated library environment", *Cataloging & Classification Quarterly*, 14 (1) 1996, 33-48.

MARKEY DRABENSTOTT, K. (1986) "Users and the online catalog : subject access problems", *The impact of online catalogs*. Edited by J. R. Matthews. New York; London : Neal-Schuman Publishers, 1986, 35-70.

MARKEY DRABENSTOTT, K. (1987) "Searching and browsing the Dewey Decimal Classification in an online catalog", *Cataloging & Classification Quarterly*, 7 (3) 1987, 37-68.

MARKEY DRABENSTOTT, K. (1990) "Experiences with online catalogs in the USA using classification system as a subject searching tool", *Tools for knowledge organization and the human interface : proceedings of the 1st International ISKO Conference, Darmstadt, 14-17 August 1990*. Edited by R. Fugmann. Frankfurt/Main : Indeks Verlag, 1990. (Advances in knowledge organization 1), 35-46.

MARKEY DRABENSTOTT, K. (1992) "The need for machine-readable authority records for topical subdivision", *Information Technology and Libraries*, 11 (2) 1992, 91-104.

MARKEY DRABENSTOTT, K. (1996) "Authority control in digital libraries", *Authority Control in the 21st Century : an invitational conference, Dublin, OH, March 31 - April 1, 1996*. http://digitalarchive.oclc.org/da/ViewObject.jsp?objid=0000003520&reqid=354.

MARKEY DRABENSTOTT, K. (1996a) "Classification to the rescue : handling the problems of too many and too few retrievals", *Knowledge organization and change : proceedings of the Fourth International ISKO Conference, Washington, DC, 15-18 July 1996*. Edited by R. Green. Frankfurt/Main : Indeks Verlag, 1996b. (Advances in knowledge organization 5). 107-136.

MARKEY DRABENSTOTT, K.; BURMAN, C. M.; WELLER, M. S. (1996) "Enhancing a new design for subject access to online catalogs", *Library Hi Tech*, 14 (1) 1996, 87-109. http://www-personal.si.umich.edu/~ylime/enhance/ENHANCE.pdf.

MARKEY DRABENSTOTT, K.; RIESTER, L. C.; DEDE, B. A. (1992) "Shelflist using expert systems", *Classification research for knowledge representation and organization : proceedings of the 5th International Study Conference on Classification Research, Toronto, Canada, 24-28 June 1991*. Edited by N. J. Williamson, M. Hudon. Amsterdam : Elsevier Science Publishers; The Hague : FID, 1992, 199-207.

MARKEY DRABENSTOTT, K.; WELLER, M. S. (1996) "A comparative approach to system evaluation : delegating control of retrieval tests on an experimental online system", *ASIS Annual Conference Proceedings, October 19-24, 1996.* http://www.asis.org/annual-96/ElectronicProceedings/Drabenstott/.

MARKEY DRABENSTOTT, K.; WELLER, M. S. (1996a) "Failure analysis of subject searches in a test of new design for subject access to online catalogs", *Journal of the American Society for Information Science*, 47 (7) 1996, 519-537.

Master Reference File : manual (2003). [UDCC documentation]. Version February 2003. The Hague : UDC Consortium, 2003. http://www.udcc.org/Manual%20version%20Feb%202003.pdf.

MAYNE, A. J. (1997) "Evolving user-oriented classifications", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 152-156.

MERHOLZ, P. (2001) "Innovation in classification", [personal blog archive] 2001. http://peterme.com/archives/00000063.html.

MICCO, M. (1996) "Subject Authority Control in the World of Internet", *LIBRES - Library Information Science Research Electronic Journal*, 6 (3) 1996. http://libres.curtin.edu.au/libre6n3/micco_2.htm.

MIKSA, F. L. (1992) "The concepts of the universe of knowledge and the purpose of LIS classification", *Classification research for knowledge representation and organization : proceedings of the 5th International Study Conference on Classification Research, Toronto, Canada, 24-28 June 1991.* Edited by N. J. Williamson, M. Hudon, Amsterdam : Elsevier Science Publishers; The Hague : FID, 1992, 101-121.

MIKSA, F. L. (1998) The DDC, the universe of knowledge, and the post-modern library. Albany, NY : Forest Press, 1998.

MILLER, P. (1996) "Metadata for the masses", *Ariadne* : The Web Version, September 1996. http://www.ariadne.ac.uk/issue5/metadata-masses/.

MILLER, P. (2000) "Interoperability : what is it and why should I want it?", *Ariadne*, 24 2000. http://www.ariadne.ac.uk/issue24/interoperability/intro.html.

MILLER, U.; TEITELBAUM, R. (2002) "Pre-coordination and post-coordination", *Knowledge Organization*, 29 (2) 2002, 87-93.

MILLS, J. (1960) A modern outline of library classification. London : Chapman & Hall, 1960.

MILLS, J. (1997) "Introductory address : [at the Sixth International Study Conference on Classification Research, London, 16-18 June 1997]", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997.* The Hague : FID, 1997. (FID 716), 1-10.

MILLS, J.; BROUGHTON, V. (1977) "Introduction and auxiliary schedules", *Bliss Bibliographic Classification*. 2nd ed. London : Bowker Saur, 1977, 29-47.

MILSTEAD, J. L. (1994) "Needs for research in indexing", *Journal of the American Society for Information Science*, 45 (8) 1994, 577-582. [Reprinted in: From classification to "knowledge organisation". Edited by A. Gilchrist. The Hague : FID, 1997. (FID 714), 151-159.]

MITCHELL, J. S. (1996) "The Dewey Decimal Classification at 120 : edition 21 and beyond", *Knowledge organization and change : proceedings of 4th International ISKO Conference, Washington DC, 15-18 July.* Edited by R. Green. Frankfurt/Main : Indeks Verlag, 1996, 15-18.

MITCHELL, J. S. (1997) "Challenges facing classification systems : a Dewey case study", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 85-89.

MITCHELL, J. S. (2000) "The Dewey Decimal Classification", *The future of classification*. Edited by R. Marcella, A. Maltby. Aldershot : Gower, 2000, 81-92.

MITCHELL, J. S.; VIZINE-GOETZ, D. (2000) "DDC taxonomy server", *Dynamism and stability of knowledge organization : proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada*. Edited by C. Beghtol, L. C. Howarth, N. J. Williamson. Würzburg : Ergon Verlag, 2000.

(Advances in knowledge organization 7), 282-287.

MIZZARO, S. (1997) "Relevance : the whole history", *Journal of the American Society for Information Science,* 48 (9) 1997, 810-832.

MOEN, W. E. (2001) "Mapping the interoperability landscape for networked information retrieval : workshop report", IITA DL Workshop, 2001. http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/main.html.

MOLHOLT, P. (2001) "The Art and Architecture Thesaurus : controlling relationships through rules and structure", *Relationships in the organization of knowledge*. Edited by C. A. Bean and R. Green. Dordrecht; Boston; London : Kluwer, 2001, 153-170.

MÖLLER, G. et al. (1999) "Automatic classification of the World Wide Web using Universal Decimal Classification", *23rd International Online Information Meeting, London, 7-9 December 1999 : proceedings*. Edited by B. McKenna. Oxford : Learned Information Europe, 1999, 231-237.

MULVANY, N. (1996) "Comments on Steinberg's article about web indexing (Steve G.Steinberg "Seek and ye shall find (maybe)", *Wired,* May 1996". http://www.mnsinc.com/curr/nanindex.htm.

MULVANY, N.; MILSTEAD, J. (1994) "Indexicon, the only fully automatic indexer : a review", *Key Words*, 2 (5) 1994, 17-23.
http://www.bayside-indexing.com/idxcon.htm.

MUSTAFA ELHADI, W. (1991) "The contribution of terminology to the theoretical conception of classification languages and document indexing", *Tools for knowledge organization and the human interface : proceedings of the 1st International ISKO Conference, Darmstadt, 14-17 August 1990*. Edited by R. Fugmann. Frankfurt/Main : Indeks Verlag, 1991. (Advances in knowledge organization 2), 98-107.

MYLOPOLUS, J.; JURISICA, I.; YU, E. (1998) "Computational mechanisms for knowledge organization", *Structures and relations in knowledge organization : proceedings of the Fifth International ISKO Conference, Lille, 25-29 August 1998*. Edited by W. Mustafa Elhadi, J. Maniez, S. Pollitt. Würzburg : Ergon Verlag, 1998. (Advances in knowledge organization 6), 125-132.

NAKAMURA, Y.; ISHIKAWA, T. (1990) "Expert systems for automatic UDC number assignment", *The UDC : essays for a new decade*. Edited by A. Gilchrist, D. Strachan. London : Aslib, 1990, 67-84.

NEELAMEGHAN, A. (1976) "Teoretska osnova UDK - zahtevi i formulacija [A theoretical foundation for UDC : its need and formulation]", *Proceedings of the international symposium "UDC in Relation to Other Indexing Languages" held in Herceg Novi, Yugoslavia, June 28-July 1, 1971.* Beograd : Jugoslovenski centar za tehnicku i naucnu dokumentaciju; The Hague : FID, 1976, 137-173.

NEELAMEGHAN, A. (1997) "A knowledge base and hypertext facility to support spiritual, moral and ethical studies", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 186-195.

NEELAMEGHAN, A.; CHOWDHURY, G. G.; CHOWDHURY, S. (1995) "VOCON : vocabulary control online in MicroISIS databases", *Knowledge Organization,* 22 (1) 1995, 18-22.

NEUROTH, H.; KOCH, T. (2001) "Cross-browsing and cross-searching in a distributed network of subject gateways : architecture, data model and classification" *ELAG 2001, Integrating Heterogeneous Resources, 25 Library Systems Seminar, Prague, 6-8 June 2001*.
http://www.stk.cz/elag2001/Papers/HeikeNeuroth/HeikeNeuroth.ppt.

NEWCOMBE, D. (1972) "SRC : unverified assumptions", *Aslib Proceedings*, 24 (10) 1972, 587-590.

NEWTON, R. (2000) "Information technology and new directions", *The future of* *classification*. Edited by R. Marcella, A. Maltby. Aldershot : Gower, 2000, 43-57.

NICHOLSON, D.; WAKE, S. (2003) "HILT : subject retrieval in a distributed environment", *Subject retrieval in a networked environment : proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001*. Edited by I. C. McIlwaine. München : K. G. Saur, 2003. (UBCIM Publications - New Series 25), 61-67.

NICHOLSON, S. (2000) "A proposal for categorization and nomenclature for Web Search Tools", *Internet searching and indexing : the subject approach*. Edited by A. R. Thomas, J. R. Shearer. New York etc : The Haworth Information Press, 2000, 9-28.

NILBE, S. (1997) "Universaalse kümnendliigituse (UDK) rahvusvahelisest levikust [The worldwide use of the Universal Decimal Classification]", Tartu Ülikooli Raamatukogu Töid = *Publications of Tartu University Library*, *1997*. http://www.utlib.ee/ee/publikatsioonid/1997/ryt/sirje.html.

NOY, N. F.; MCGUINNESS, D. L. (2001) "Ontology development 101 : a guide to creating your first ontology", Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.
http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html.

NWALI, L. O. (1992) "Indexing services in an academic library in Nigeria", *International Cataloguing and Bibliographic Control*, 21 (3) 1992, 42-44.

ÖHMAN, E.; OLIVECRONA, Ch. (1976) "Neki notacijski hijerarhijski i sintaktički problemi u vezi sa usaglašavanjem UDK i tezaurusa [Some notational, hierarchic, and syntactic problems in connection with concordances between UDC and thesauri]", *Proceedings of the international symposium "UDC in Relation to Other Indexing Languages" held in Herceg Novi, Yugoslavia, June 28-July 1, 1971*. Beograd : Jugoslovenski centar za tehnicku i naucnu dokumentaciju; The Hague :

FID, 1976, 175-214.

OVERFIELD, C. (2000) "MRF Scanner Utility : guideline". Documentation for the MRF Scanner program [UDC Consortium internal document], London, 2000.

PANIGRAHI, P.; PRASAD, A. R. D. (1997) "An inference engine for time isolates in Colon Classification", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 196-199.

PARAMESWARAN, M. (1990) "Chain procedure and Dewey Decimal Classification", *Tools for knowledge organization and the human interface : proceedings of the 1st International ISKO Conference, Darmstadt, 14-17 August 1990*. Edited by R. Fugmann. Frankfurt/Main : Indeks Verlag, 1990. (Advances in knowledge organization 1), 250-258.

PEDERSEN, N. (2004) "UDC in Denmark". 2004-10-04. Private e-mail message.

PEKKARINEN, P. (1997) "On 'new' Ranganathanian laws for structuring the content of the virtual library", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 144-146.

PEPPER, S.; GARSHOL, L. M. (2002) "Lessons on applying Topic Maps", [paper presented at] *XML Conference and Exposition, Baltimore MD, December 8-13 2002.* http://www.ontopia.net/topicmaps/materials/xmlconf.htmlXML-Europe-2002.

PERREAULT, J. M. (1969) Towards a theory for UDC : essays aimed at structural understanding and operational improvement. London : Clive Bingley, 1969.

PERREAULT, J. M. (1969a) "On the colon in UDC and on the Boolean operators", *Proceedings of the First Seminar on UDC in a Mechanized Retrieval System conducted by R. R. Freeman and P. Atherton, Copenhagen, 2th[sic!]-6th September, 1968*. Copenhagen : Danish Centre for Documentation, 1969. (FID/CR Report no 9), 38-41.

PERREAULT, J. M. (1976) "Pojam 'glavne klase' i mogućnost za poboljšanje UDK [The concept 'main class' and its implication for structural improvement of UDC]", *Proceedings of the international symposium "UDC in Relation to Other Indexing Languages" held in Herceg Novi, Yugoslavia, June 28-July 1, 1971*. Beograd *:* Jugoslovenski centar za tehnicku i naucnu dokumentaciju; The Hague : FID, 1976, 215-229.

PERREAULT, J. M. (1978) "Commentaries to the D. Austin paper", *A conceptual approach to the organization of machine-held files for subject retrieval", Conceptual basis of the classification of knowledge : proceedings of the Ottawa Conference, October 1st to 5th 1971*. Edited by J. A. Wojciechowski. New York; München; Paris : K.G. Saur, 1978, 399-403.

PERREAULT, J. M. (1994) "Categories and relators", *Knowledge Organization,* 21 (4) 1994, 189-198.

PETERS, T. A. (1991) The online catalog : a critical examination of public use. Jefferson, NC; London : McFarland, 1991.

PIKA, J. (2002) "Anwendung der UDK in NEBIS in der Schweiz : ein Ausblick", [talk given at the seminar] *Jahrestagung der Gesellschaft für Klassifikation (GfKl), Universität Mannheim, 23. Juli 2002*.
http://www.ethbib.ethz.ch/pub/nebisUDK.pdf.

POLI, R. (1996) "Ontology for knowledge organization", *Knowledge organization and change : proceedings of the Fourth International ISKO Conference, Washington, DC, 15-18 July 1996*. Edited by R. Green. Frankfurt/Main : Indeks Verlag, 1996. (Advances in knowledge organization 5), 313-319.

POLLITT, S. (1997) "Interactive information retrieval based on faceted classification using views", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research,* *London, 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 51-56.

POLLITT, S. (1997a) "The key role of classification and indexing in view-based searching", *63rd IFLA General Conference Copenhagen, Denmark, August 31 - September 5, 1997*. http://www.ifla.org/IV/ifla63/63polst.pdf

POLLITT, S. (1998) "The application of Dewey Classification in a view-based searching OPAC", *Structures and relations in knowledge organization : proceedings of the Fifth International ISKO Conference, Lille, 25-29 August 1998*. Edited by W. Mustafa Elhadi, J. Maniez, S. Pollitt. Würzburg : Ergon Verlag, 1998. (Advances in knowledge organization 6), 176-183.

POLLITT, S.; TINKER, A. J. (2000) "Enhanced view-based searching through the decomposition of Dewey Decimal Classification Codes", *Dynamism and stability of knowledge organization : proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada*. Edited by C. Beghtol, L. C. Howarth, N. J. Williamson. Würzburg : Ergon Verlag, 2000. (Advances in knowledge organization 7), 288-294.

PORTER, M.; GALPIN, V. (1988) "Relevance feedback in a public access catalogue for a research library : MUSCAT at the Scott Polar Research Institute Library", *Program,*. 22 (1) 1988, 1-20.

POPPER, K. R. (1979) Objective knowledge. Oxford : Oxford University Press, 1979.

POWELL, A. (1997) "Dublin Core Management", *Ariadne* : The Web Version, July 1997. http://www.ariadne.ac.uk/issue10/dublin/.

POWELL, A. (1998) "Identification - Related resources", BIBLINK project, Work package 2, 1998. http://hosted.ukoln.ac.uk/biblink/wp2/links.html.

POWELL, A. (2003) DC and external authority data. 2003-03-05. Message to the mailing list dc-general@jiscmail.ac.uk.

http://www.jiscmail.ac.uk/lists/dc-general.html.

POWELL, A. et al. (2001) "Namespace policy for the Dublin Core Metadata Initiative". DCMI, 2000.
http://dublincore.org/documents/2001/09/17/dcmi-namespace/.

POWELL, A.; JOHNSTON, P. (2002) "Guidelines for implementing Dublin Core in XML". DCMI, 2002.
http://www.ukoln.ac.uk/metadata/resources/dc/dc-xml-guidelines/.

PRASAD, A. R. D. (1996) "PROMETHEUS : an automatic indexing system", *Knowledge organization and change : proceedings of the Fourth International ISKO Conference, 15-18 July 1996, Washington DC*. Edited by R. Green. Frankfurt/Main : Indeks Verlag, 1996. (Advances in knowledge organization 5), 329-335.

Principles of the Universal Decimal Classification and rules for its revision and publications = Principes de la Classification Decimale Universelle et regles pour sa revision et sa publications = Grundsätze der Unversellen Dezimalklassifikation (DK) und Regeln fur ihre Revision und Veröffentlichung. The Hague : FID, 1981.

PRISS, U. (1998) "A graphical interface for conceptually navigating faceted thesaurus", *Structures and relations in knowledge organization : proceedings of the Fifth International ISKO Conference, Lille, 25-29 August 1998*. Edited by W. Mustafa Elhadi, J. Maniez, S. Pollitt. Würzburg : Ergon Verlag, 1998. (Advances in knowledge organization 6), 184-190.

Proceedings of the International Symposium "UDC in Relation to other Indexing Languages", Herceg Novi, Yugoslavia, June 28-July 1, 1971" (1976), Herceg Novi, Yugoslavia, June 28-July 1, 1971. (1971). Beograd : Jugoslovenski centar za tehnicku i naucnu dokumentaciju; The Hague : FID, 1976.

Published Subjects : introduction and basic requirements : [status : committee specification]. (2004) OASIS [The Organization for the Advancement of Structured Information Standards] Published Subjects Technical Committee Recommendation,

2003-06-24. Edited by S. Pepper.
http://www.oasis-open.org/committees/download.php/3050/pubsubj-pt1-1.02-cs.pdf.

QIYU, Z.; XIANGSHENG, L; DONGBO, W. (1996) "Contemporary classification systems and thesauri in China", *62nd IFLA General Conference, August 25-31, 199 : conference proceedings.* http://www.ifla.org/IV/ifla62/62-qiyz.htm.

QUINN, B. (1994) "Recent theoretical approaches in classification and indexing", *Knowledge Organization*, 21 (3) 1994, 140-147.

RANGANATHAN, S. R. (1961) "Library classification on the march", *The Sayers Memorial Volume : essays in librarianship in memory of William Charles Berwick Sayers.* Edited by D. J. Foskett, B. I. Palmer. London : The Library Association, 1961, 72-95.

RANGANATHAN, S. R. (1962) Elements of library classification. 3rd ed. London : Asia Publishing House, 1962.

RANGANATHAN, S. R. (1967) Prolegomena to library classification. 3rd ed. London : Asia Publishing House, 1967.

RAYWARD, B. W. (1994) "The International Federation for Information and Documentation (FID)", *Encyclopedia of Library History*. Edited by W. A. Wiegand, D. G. Davis. New York : Garland Press, 1994, 290-294.
http://alexia.lis.uiuc.edu/~wrayward/otlet/FIDHIST2.htm.

Renardus - the clever route to information : project summary. (2000) Renardus project, 2000. http://www.renardus.org/.

Research project to automate cataloging of Internet resources seeks participants worldwide. (1998) OCLC News Releases, 1998.
http://www.oclc.org/oclc/press/980928.htm.

Resource Description Framework (RDF). (2004) World Wide Web Consortium

(W3C), 2004. http://www.w3.org/RDF/.

Resource discovery - a definition. (1995) *DTSC Symposium 1995 - position paper*. http://archive.dstc.edu.au/RDU/RD-Defn.

RIESTHUIS, G. J. A. (1994) "Conversion from precoordinated classified catalogues to postcoordinated keyword catalogues", *Knowledge organization and quality management : proceedings of the Third International ISKO Conference, 20-24 June 1994, Copenhagen, Denmark.* Edited by H. Albrechtsen, S. Oernager. Frankfurt/Main : Indeks Verlag, 1994. (Advances in knowledge organization 4), 398-405.

RIESTHUIS, G. J. A. (1997) "Decomposition of complex UDC notation", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 139-143.

RIESTHUIS, G. J. A. (1998) "Decomposition of UDC-numbers and the text of the UDC Master Reference File", *Structures and relations in knowledge organization : proceedings of the Fifth International ISKO Conference, Lille, 25-29 August 1998*. Edited by W. Mustafa Elhadi, J. Maniez, S. Pollitt. Würzburg : Ergon Verlag, 1998. (Advances in knowledge organization 6), 221-228.

RIESTHUIS, G. J. A. (1998a) "The UDC Master Reference File", *Extensions & Corrections to the UDC*, 20 1998, 21-26.

RIESTHUIS, G. J. A. (1998b) Zoeken met woorden. [doctoral thesis]. Amsterdam : Leerstoelgroep Boek-, Archief- en Informatiewetenschap, 1998. (BBI-reeks nr 6).

RIESTHUIS, G. J. A. (1999) "Searching with words : re-use of subject indexing", *Extensions & Corrections to the UDC*, 21 1999, 24-32.

RIESTHUIS, G. J. A. (2000) "Multilingual subject access and the Guidelines for the establishment and development for multilingual thesauri", *Dynamism and stability of* *knowledge organization : proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada*. Edited by C. Beghtol, L. C. Howarth, N. J. Williamson. Würzburg : Ergon Verlag, 2000. (Advances in knowledge organization 7), 131-135.

RIESTHUIS, G. J. A. (2001) "Some thoughts about the format of the Master Reference File database. *Extensions & Corrections to the UDC*, 23 2001, 15-22.

RIESTHUIS, G. J. A. (2002) "A Revised Format for the Master Reference File Database", *Extensions & Corrections to the UDC*, 24 2002, 11-15.

RIESTHUIS, G. J. A. (2003) "A Revised Format for the Master Reference File Database, *Extensions & Corrections to the UDC*, 25 2003, 11-18.

RIESTHUIS, G. J. A. (2003a) "Information languages and multilingual subject access", *Subject retrieval in a networked environment : proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001*. Edited by I. C. McIlwaine. München : K. G. Saur, 2003. (UBCIM Publications - New Series 25), 11-17.

RIESTHUIS, G. J. A.; BLIEDUNG, S. (1990) "Thesaurification of UDC : preliminary report", *The UDC : essays for a new decade*. Edited by A. Gilchrist, D. Strachan. London : Aslib, 1990, 85-97.

RIESTHUIS, G. J. A.; BLIEDUNG, S. (1991) "Thesaurification of the UDC", *Tools for knowledge organization and the human interface : proceedings : 1st International ISKO Conference, Darmstadt, 14-17 August*. Edited by R. Fugmann. Frankfurt/Main : Indeks Verlag, 1991. (Advances in knowledge organization 2), 109-117.

RIGBY, M. (1971) "The role of the UDC in automated information and data systems", *Proceedings of the Second Seminar on UDC in Mechanized Information Systems conducted by R. R. Freeman, Frankfurt, 1st -5th June, 1970*. Copenhagen : Danish Centre for Documentation, 1971. (FID/CR Report no 11), 182-192 .

RIGBY, M. (1974) Computers and the UDC : a decade of progress 1963-1973. The Hague : FID, 1974. (FID 523).

RIGBY, M. (1976) "Inverzni UDK - indeks kao korak dalje ka stvaranju i održavanju uprošćenog i efikasnog tezaurusa [An inverted UDC index as a step toward simplified and effective thesaurus development and maintenance]", *Proceedings of the international symposium "UDC in Relation to Other Indexing Languages" held in Herceg Novi, Yugoslavia, June 28-July 1, 1971*. Beograd *:* Jugoslovenski centar za tehnicku i naucnu dokumentaciju; The Hague : FID, 1976, 231-236.

RIGBY, M. (1981) Automation and the UDC : 1948 - 1980. 2nd ed. The Hague : FID, 1981. (FID 565).

RIJSBERGEN, C. J. van. (1979) Information retrieval. 2nd ed. London : Butterworths, 1979.

RINNE, B. (1990) "ELECON system : economics database", *The UDC : essays for a new decade*. Edited by A. Gilchrist, D. Strachan. London : Aslib, 1990, 55-66.

ROBINSON, G. (2001) "Fixed citation order?", *Extensions & Corrections to the UDC*, 23 2001, 8.

ROBINSON, G. (2003) "Citation order in UDC", *Extensions & Corrections to the UDC*, 25 2003, 19-27.

ROMARY, L ; CAMPENHOUDT, M. van. (2001) "Normalisation des échanges de données en terminologie : le cas des relations dites 'conceptuelles'", [paper presented at] *Conférence TIA-2001, Nancy, 3 et 4 mai 2001*.
http://www.termisti.refer.org/tia4.pdf.

ROWLEY, J. (1994) "The controlled versus natural indexing languages debate revisited : a perspective on information retrieval practice and research", *Journal of Information Science*, 20 (2) 1994, 108-119.

SALEM, S. (2004) "UDC in Arabic". 2004-08-13. Private e-mail message.

SALTON, G. (1989) Automatic text processing : the transformation, analysis, and retrieval of information by computer. Reading [etc.] : Addison Wesley Publishing Company, 1989.

SANTORO, M. (1995) "Ripensare la CDU : per una riflessione sulla storia, il ruolo e le prospettive della Classificazione Decimale Universale ", *Biblioteche oggi*, 13 (8) 1995, 48-57. http://www.aib.it/aib/contr/santoro1.htm.

SANTORO, M. (1996) "International exchange languages : the case of the Universal Decimal Classification", *Extensions & Corrections to the UDC,* 18 1996, 81-91.

SAPIIE, J. (1995) "Reader-interest classification : the user-friendly schemes", *Cataloging & Classification Quarterly*, 19 (3-4) 1995, 143-155.

SARACEVIC, T. (1996) "Relevance reconsidered : information science : integration in perspectives", *Proceedings CoLIS2 , Second International Conference on Conception of Library and Information Science : integration in perspective, October 13-16, 1996*. Edited by P. Ingwersen, N. Ole Pors. Copenhagen : The Royal School of Librarianship, 1996. 201-218.

SATIJA, M. P. (1993) "A deserted bastion : classification and indexing in India a state-of-the-art", *International Cataloguing and Bibliographic Control*, 22 (3) 1993, 51-54.

SATIJA, M. P. (2000) "Sources for investigating the development of bibliographical classification", *The future of classification.* Edited by R. Marcella, A. Maltby. Aldershot : Gower, 2000, 129-140.

SAYE, J. D. (1990) "The Library of Congress Classification System in an online environment : a reaction", *Cataloging & Classification Quarterly*, 11 (1) 1990, 27-35.

SCHALLIER, W. (2004) "On the razor's edge : between local and overall needs in knowledge organization", *Knowledge Organization and the Global Information Society : proceedings of the 8th ISKO Conference, London 13-16 July 2004*. Edited by I. C. McIlwaine. Würzburg : Ergon Verlag, 2004. (Advances in knowledge organization 9), 269-274.

SCHALLIER, W. (2004a) "What a subject search interface can do", *Extensions & Corrections to the UDC*, 26 2004, 18-28.

The SCHEMAS Registry. (2000) Forum for Metadata Schemas Implementers, 2000. http://www.schemas-forum.org/registry/.

SCHNEIDER, J. H. (1979) "Autoclass : a computer system for facilitating the creation and updating of hierarchical classifications", *Ordering systems for global information networks : proceedings of the Third International Study Conference on Classification Research held at Bombay, India, during 6-11 January 1975*. Edited by A. Neelameghan, Bangalore : DRTC : FID/CR and Sarada Ranganathan Endowment for Library Science, 1979. (FID 533), 269-274.

SCHUEGRAF, E. J. (1997) "Classification as an aid to automatic indexing", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 67-71.

ŚCIBOR, E. (1976) "Neka zapažanja o uporednoj efikasnosti UDK i deskriptorskih jezika [Some remarks on the comparative efficiency of UDC and descriptor languages]", *Proceedings of the international symposium "UDC in Relation to Other Indexing Languages" held in Herceg Novi, Yugoslavia, June 28-July 1, 1971.* Beograd *:* Jugoslovenski centar za tehnicku i naucnu dokumentaciju; The Hague : FID, 1976, .245-251.

ŚCIBOR, E. (1996) "Activities of the ISTEI in the field of the UDC", *Extensions & Corrections to the UDC*, 18 1996, 93-97.

ŚCIBOR, E. (1997) "UDC as a tool for information retrieval", *Knowledge Organization for Information Retrieval : proceedings of the Sixth International Study Conference on Classification Research, London 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 200-204.

ŚCIBOR, E.; SHCHERBINA-SAMOJLOVA, I. S. (1990) "A strategic approach to revising the UDC", *The UDC : essays for a new decade*. Edited by A. Gilchrist, D. Strachan. London : Aslib, 1990, 11-18. (reprint from : Information, knowledge, evolution : proceedings of the forty-fourth FID congress held in Helsinki, Finland, 28 August-1 September, 1988. Edited by S. Koskiala, R. Launo. Amsterdam; Oxford : North-Holland, 1989 (FID publication; 675).

ŚCIBOR, E.; TOMASIK-BECK, J. (1994) "On the establishment of concordances between indexing languages of universal disciplinary scope", *Knowledge Organization*, 21 (4) 1994, 203-212.

Second Seminar on UDC in Mechanized Information Systems (1971) conducted by R. R. Freeman, Frankfurt, 1st-5th June, 1970 : proceedings. Edited by R. Molgaard-Hansen, M. Westring-Nielsen. Copenhagen : Danish Centre for Documentation, 1971. (FID/CR Report no 11).

Semantic Web. (2004) World Wide Web Consortium (W3C), 2004.
http://www.w3.org/2001/sw/.

SHAFER, K. (1996) "A brief introduction to Scorpion". OCLC, 1996.
http://purl.oclc.org/scorpion/bintro.html (not available on the Web anymore).

SHAFER, K. (1997) "Scorpion helps catalog the Web", *Bulletin of the American Society for Information Science*, 24 (1) 1997. http://www.asis.org/Bulletin/Oct-97/shafer.htm.

SHAFER, K. (1998) "Evaluating Scorpion results", OCLC, 1998.
http://orc.rsch.oclc.org:6109/eval_sc.htm.

SHAFER, K; THOMPSON, R. (1998) "Scorpion : SMART weighting scheme". OCLC, 1998. http://orc.rscn.oclc.org:6109/smart_weight.html.

SHERMAN, C. (1999) "The invisible Web", *Your About.com guide to Web Search*. http://websearch.about.com/library/weekly/aa061199.htm.

SHIRI, A. A.; REVIE, C.; CHOWDHURY, G. (2002) "Thesaurus-assisted search term selection and query expansion : a review of user-centred studies", *Knowledge Organization*, 29 (1) 2002, 1-17.

SHNEIDERMAN, B.; BYRD, D.; CROFT, B. (1997) "Clarifying search : a user-interface framework for text searches", *D-lib Magazine*, January 1997. http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html.

SHOHAM, S. (2000) Library classification and browsing. Brighton; Portland, OR : Academic Press, 2000.

SIGEL, A. (2004) "Wissensorganisation, Topic Maps und Ontology Engineering : die Verbindung bewährter Begriffsstrukturen mit aktueller XML-Technologie", [presented at 7. Deutschen ISKO Tagung in Berlin, March 2004 [to be published in Tagungsband der 7. Deutschen ISKO-Tagung, Würzburg : Ergon Verlag]. http://kpeer.wim.uni-koeln.de/~sigel/veroeff/ISKO-2001/ISKO2001_Berlin_SIGEL_v24.pdf.

SILVA, A.; SANTOS, M.; LOPES, M. (1989) "SIPORbase : um novo sistema de indexação", *Revista da Biblioteca Nacional*, 4 (1) l989, 117-129.

SKRETAS, G. (2004) "UDC in Greek". 2004-09-13. Private e-mail message.

SLAVIC, A. (2001) "Predmetni pristup informacijama na Internetu i knjižnična klasifikacija [Subject approach to information on the Internet and library classification]", *Vjesnik bibliotekara Hrvatske*, 44 (1-4) 2001, 67-81.

SLAVIC, A. (2001a) "UDK i tablice drugi veliki klasifikacijski sustavi dostupni na

Web-u [UDC and other large classification systems available on the Web]", *Vjesnik bibliotekara Hrvatske*, 44 (1-4) 2001, 95-109.

SLAVIC, A. (2003) "General library classification in learning material metadata : the application in IMS/LOM and DCMES metadata schemas", *Subject retrieval in a networked environment : proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001*. Edited by I. C. McIlwaine. München : K. G. Saur, 2003. (UBCIM Publications - New Series 25), 97-104.

SLAVIC, A. (2003a) "UDC implementation : from library shelves to a structured indexing language", [paper presented at] *69th IFLA General Conference and Council, World Library and Information Congress, 1-9 August 2003, Berlin*.
http://www.ifla.org/IV/ifla69/papers/032e-Slavic.pdf.
[Published in: *International Cataloguing and Bibliographic Control*, 33 (3) 2004, 60-65.]

SLAVIC, A. (2003b) FATHUM classification database tool.
http://www.ucl.ac.uk/fatks/database.htm.

SLAVIC, A. (2003c) "Classification structure and data model (Work Package 3 : Deliverable 4)". AHRB - Innovation Awards Scheme Project : Facet Analytical Theory in Knowledge Structure for Humanities. Final version. September 2003.
http://www.ucl.ac.uk/fatks/restricted/D4-Classification_structure.doc [project internal document].

SLAVIC, A. (2004) "UDC translations : a 2004 survey report and bibliography", *Extensions & Corrections to the UDC*, 26 2004, 58-80.

SLAVIC, A.; BAIGET, C. (2001) "Using Dublin Core in educational material : some practical considerations based on the EASEL experience", *Vine*, 31 (4) 2001, 74-85.

SLAVIC, A.; CORDEIRO, M. I. (2004) "Core requirements for automation of analytico-synthetic classifications", *Knowledge Organization and the Global Information Society : proceedings of the 8th ISKO Conference, London 13-16 July*

2004. Edited by I. C. McIlwaine. Würzburg : Ergon Verlag, 2004. (Advances in knowledge organization 9), 187-192.

SLAVIC, A.; CORDEIRO, M. I. (2004a) "Sharing and re-use of classification systems : the need for a common data model", *Signum*, 8 2004, 19-24.

SLAVIC, A.; TURKULIN, B. (1998) "Prevajanje klasifikacijskih oznak v naravni jezik [Library classification on-line : translation of notation into natural language]", *Knjižnica*, 42 (1) 1998, 45-63.

SOERGEL, D. (1994) "Indexing and retrieval performance : the logical evidence", *Journal of the American Society for Information Science*, 45 (8) 1994, 589-599.

SOERGEL, D. (1996) "SemWeb : proposal for an open, multifunctional, multilingual, system for integrated access to knowledge : about concepts and terminology", *Knowledge organization and change : proceedings of the Fourth International ISKO Conference, 15-18 July 1996, Washington DC*. Edited by R. Green. Frankfurt/Main : Indeks Verlag, 1996. (Advances in knowledge organization 5), 165-173.

SOERGEL, D. (1999) "The rise of ontologies or the reinvention of classification", *Journal of the American Society for Information Science*, 50 (12) 1999, 1119-1120.

SOERGEL, D. (2001) "The representation of Knowledge Organization Structure (KOS) data : a multiplicity of standards", [paper presented at] *JCDL NKOS Workshop, VA, 2001*. http://www.dsoergel.com/cv/B75.pdf.

SOSIG Guidelines for assigning keywords : section editors workshop. (2004) SOSIG, ILRT, University of Bristol, June 2004 [internal document].

SOWA, J. F. (2000) Knowledge representation : logical, philosophical and computational foundations. Pacific Grove,CA [etc.] : Brooks/Cole, 2000.

SPARCK JONES, K. (1970) "Some thoughts on classification for retrieval", *Journal* *of Documentation*, 29 (2) 1970, 89-101.

SPARCK JONES, K. (1995) "Reflections on TREC", *Information Processing & Management*, 31 (3) 1995, 291-314. [Reprinted in: From classification to "knowledge organisation", Dorking revisited or "past is prelude". Edited by A. Gilchrist. The Hague : FID, 1997. (FID 714), 101-124.]

SPINK, A.; SARACEVIC, T. (1997) "Interaction in information retrieval : selection and effectiveness of search terms", *Journal of the American Society for Information Science*, 48 (8) 1997, 741-761.

SPITERI, L. F. (1995) "The Classification Research Group and the theory of integrative levels". http://edfu.lis.uiuc.edu/review/summer1995/spiteri.html.

SRIMURUGAN, A. (2001) "An expert system for generation of UDC class numbers : an investigation", *Extensions & Corrections to the UDC*, 23 2001, 25-30.

STAR, S. L. (1998) "Grounded classification : grounded theory and faceted classification", *Library Trends*, 47 (2) 1998, 218-232.

STOKLASOVÁ, B. (2003) "Short survey of subject gateways (SG) activity", IFLA [paper presented at] *69th IFLA General Conference and Council, World Library and Information Congress, 1-9 August 2003, Berlin*.
http://www.ifla.org/IV/ifla69/papers/152e-Stoklasova.pdf.

STOKLASOVÁ, B.; BALIKOVÁ, M.; CELBOVÁ, L. (2003) "The relationship between subject gateways and national bibliographies in international context", *IFLA [paper presented at] 69th IFLA General Conference and Council, World Library and Information Congress, 1-9 August 2003, Berlin*.
http://www.ifla.org/IV/ifla69/papers/054e-Stoklasova_Balikova_Celbova.pdf.

STRACHAN, D. (1990) "UDC revision work in FID", *The UDC : essays for a new decade*. Edited by A. Gilchrist, D. Strachan. London : Aslib, 1990, 1-10.

STRACHAN, D. (2000) "A chance to make UDC notation more user friendly", *Extensions & Corrections to the UDC*, 22 2000, 43-46.

STRACHAN, D. (2004) UDC user survey. 2004-08-19 and 2004-08-22. Private e-mail messages.

STRACHAN, D. (2004a) UDC Translations. 2004-11-05. Private e-mail message.

STRACHAN, D.; OOMES, F. M. H. (1993) "The UDC Master Reference File (MRF)", *Extensions & Corrections to the UDC*, 15 1993. The Hague : UDC Consortium, 1993, 19-28.

STRACHAN, D.; OOMES, F. M. H. (1995) "Universal Decimal Classification update", *Cataloging and Classification Quarterly*, 19 (3-4) 1995, 119-131.

STUEART, R. D. (1971) "An analysis of the Universal Decimal Classification as a term system for nuclear science and technology", *Library Resources & Technical Services*, 15 (3) 1971, 399-411.

Subject data in the metadata record : recommendations and rationale : a report from the ALCTS/CCS/SAC/Subcommittee on Metadata and Subject Analysis. (1999) July 1999. http://www.govst.edu/users/gddcasey/sac/MetadataReprot.html; http://archive.ala.org/alcts/organization/ccs/sac/metaclassfinal.pdf.

SUBRAMANIAN, S.; SHAFER, K. (1997) "Clustering". OCLC, 1997. http://digitalarchive.oclc.org/da/ViewObject.jsp?objid=0000003409.

SUKIASYAN, E. R. (1988) "Classification practice in the USSR : current status and development trends", *International Classification*, 15 (2) 1988, 69-72.

SUKIASYAN, E. R. (1996) "Changes as a problem of classification system development", *Knowledge organization and change : proceedings of the Fourth International ISKO Conference, 15-18 July 1996, Washington DC*. Edited by R. Green. Frankfurt/Main : Indeks Verlag, 1996. (Advances in knowledge organization

5), 119-121.

SUKIASYAN, E. R. (1998) "Classification systems in their historical development : problems of typology and terminology", *Structures and relations in knowledge organization : proceedings of the Fifth International ISKO Conference, Lille, 25-29 August 1998*. Edited by W. Mustafa Elhadi, J. Maniez, S. Pollitt. Würzburg : Ergon Verlag, 1998. (Advances in knowledge organization 6), 71-78.

SUTTON, S. A. (1999) "Conceptual design and deployment of a metadata framework for educational resources on the Internet", *Journal of the American Society for Information Science*, 50 (13) 1999, 1182-1192.

SUTTON, S. A.; OH, S. G. (1997) "GEM : using metadata to enhance Internet retrieval by K-12 teachers", *Bulletin of the American Society for Information Science*, October 1997. http://www.asis.org/Bulletin/Oct-97/suttonoh.html.

SVENONIUS, E. (1982) "Indexical contexts", *Universal classification : subject analysis and ordering systems : proceeding of the 4th International Study Conference on Classification Research, 6th Annual Conference of Gesellshaft für Klassifikation, Augsburg, 28 June - 2 July 1982, Vol. 1*. Edited by I. Dahlberg. Frankfurt : Indeks Verlag, 1982, 125-138.

SVENONIUS, E. (1983) "Use of classification in online retrieval", *Library Resources and Technical Services*, 27 (1) 1983, 76-80.

SVENONIUS, E. (1986) "Unanswered questions in the design of controlled vocabularies", *Journal of the American Society for Information Science*, 37 (5) 1986, 331-340. [Reprinted in: From classification to "knowledge organisation", Dorking revisited or "past is prelude". Edited by A. Gilchrist. The Hague : FID, 1997. (FID 714), 137-147.]

SVENONIUS, E. (1992) "Classification : prospects, problems and possibilities", *Classification Research for Knowledge Representation and Organization : proceedings of the 5th International Study Conference on Classification Research,* *Toronto, Canada, June 24-28, 1991*. Edited by N. J. Williamson, M. Hudon. Amsterdam : Elsevier : FID, 1992, 5-25.

SVENONIUS, E. (1992a) "The impact of computer technology on knowledge representation", *Cognitive paradigms in knowledge organization : [proceedings of the] Second International ISKO conference, Madras, 26-28 August 1992*. Bangalore : Sarada Ranganathan Endowment for Library Science, 1992. 1-34.

SVENONIUS, E. (1995) "Precoordination or not?", *Subject indexing : principles and practices in the 90's : proceedings of the IFLA Satellite meeting held in Lisbon, Portugal, 17-18 August 1993*. Edited by Robert P. Holley et al.. München [etc.] : K. G. Saur, 1995, 231-255.

SVENONIUS, E. (1997) "Definitional approaches in the design of classification and thesauri and their implications for retrieval and for automatic classification", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 12-16.

SVENONIUS, E. (2000) The intellectual foundation of information organization. Cambridge, MA; London : The MIT Press, 2000.

SVENONIUS, E. (2004) "The epistemological foundation of knowledge representation", *Library Trends*, 52 (3) 2004, 571-587.

SVENONIUS, E.; LIU, S.; SUBRAHMANYAM, B. (1992) "Automation of chain indexing", *Classification research for knowledge representation and organization : proceedings of the 5th International Study Conference on Classification Research, Toronto, Canada, 24-28 June 1991*. Edited by N. J. Williamson, M. Hudon. Amsterdam : Elsevier; The Hague : FID, 1992, 351-364.

SWANSON, D. R. (1986) "Undiscovered public knowledge", *Library Quarterly*, 56 (2) 1986, 103-118.

SWANSON, D. R. (1987) "Two medical literatures that are logically but not bibliographically connected", *Journal of the American Society for Information Science*, 38 (4) 1987, 228-233

SWANSON, D. R. (2001) "On the fragmentation of knowledge, the connection explosion, and assembling other people's ideas : ASIST Award of Merit acceptance speech", *Bulletin of the American Society for Information Science and Technology*, 27 (3) 2001, 12-14.

St. PIERRE, M; LAPLANT, W. (1998) "Issues in crosswalking content metadata standards". National Information Standards Organization (NISO), 1998. http://www.niso.org/press/whitepapers/crsswalk.html.

Task force on UDC system development : interim report (1989). FID UDC/MB 89-18. The Hague : FID, 1989. [unpublished]

Task force for UDC system development : final report. (1990) February 1990. [FID internal document]. [unpublished]

TAYLOR, A. G. (1999) The organization of information. Englewood, CO : Libraries Unlimited, 1999.

TAYLOR, A. G. (1999a) "Authority control : where it's been and where it's going", *Authority Control : Why it Matters, November 1, 1999, Worcester, MA*. http://www.nelinet.net/edserv/conf/cataloging/cts/1999/taylor.htm.

TAYLOR, C. (1999) "An introduction to metadata", *Cybrary*, 1999. http://www.library.uq.edu.au/iad/ctmeta4.html. [revised 2003]

TENNANT, R. (2002) "Digital libraries : what to know about Web Services", *Library Journal*, 2002. http://www.libraryjournal.com/index.asp?layout=articleArchive&articleid=CA231639.

THOMPSON, R.; SHAFER, K.; VIZINE-GOETZ, D. (1997) "Evaluating Dewey concepts as a knowledge base for automatic subject assignment". OCLC, 1997. http://orc.rscn.oclc.org:6109/eval_dc.html.

TILLETT, B. (1995) "21st century authority control : what is it and how do we get there?" *The future is now : reconciling change and continuity control : proceedings of the OCLC Symposium, ALA Annual Conference, June 23, 1995.* http://digitalarchive.oclc.org/da/ViewObject.jsp?fileid=0000003587:000000094276 &reqid=228.

TILLETT, B. (2000) "Authority control on the Web", [paper presented at the] *Bicentennial Conference on Bibliographic Control for the New Millennium, Library of Congress, 2000*. http://lcweb.loc.gov/catdir/bibcontrol/tillet.html.

TOBIAS, J. (1998) "Seeking the subject", *Library Trends*, 47 (2) 1998, 209-217.

TUDHOPE, D; KOCH, T. (2004) "New Applications of Knowledge Organization Systems : introduction to a special issue" *Journal of Digital Information*, 4 (4) 2004, article no. 286. http://jodi.ecs.soton.ac.uk/Articles/v04/i04/editorial/.

UKOLN metadata pages. UKOLN, University of Bath, 2004. http://www.ukoln.ac.uk/metadata/.

The UDC : essays for a new decade. (1990) Edited by A. Gilchrist, D. Strachan. London : Aslib, 1990.

[UDC] (1993) Universal Decimal Classification : international medium edition : English text, edition 2. Part 1-2. London : British Standards Institution, 1993. (BS 1000M)

[UDC] (2003) Universal Decimal Classification : abridged edition. London : British Standards Institution, 2003.

UDC MRF (2002). UDC Master Reference File, 2002 update. The Hague : UDC Consortium, 2002.

UDC MRF (2003). UDC Master Reference File, 2003 update. The Hague : UDC Consortium, 2003.

UDC revision and publication procedure. (1968) The Hague : FID, 1968.

UNISIST study report on the feasibility of a world science information system. (1971) Paris : UNESCO, 1971.

URR, C. (1991) "Will the real hypertext please stand up?", *Computers in Libraries*, 11 (5) 1991, 46-49.

USMARC format for classification data : including guidelines for content designation. (1997). Cumulation of 1990 edition and Update No. 1 95 interfiled. Prepared by Network Development and MARC Standards Office. Washington, DC : Library of Congress, Cataloging Distribution Service, 1997.

VAN DER WALT, M. S. (1997) "The role of classification in information retrieval on the Internet : some aspects of browsing lists in search engines", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 32-35.

VAN DER WALT, M. S. (1998) "The structure of classification schemes used in Internet search engines", *Structures and relations in knowledge organization : proceedings of the Fifth International ISKO Conference, Lille, 25-29 August 1998*. Edited by W. Mustafa Elhadi, J. Maniez, S. Pollitt. Würzburg : Ergon Verlag, 1998. (Advances in knowledge organization 6), 379-387.

VAN DER WALT, M. S. (2004) Information on UDC in South Africa. 2004-08-23. Private e-mail message.

VAN DE SOMPEL, H. et al. (2004) The "info" URI scheme for information assets with identifiers in public namespaces : Network Working Group Internet-draft : Expires January 7, 2005. July 9th 2004.

http://www.ietf.org/internet-drafts/draft-vandesompel-info-uri-02.txt

VAN DIJK, P. (2003) "eXchangeable Faceted Metadata Language - XFML Core . [Last changed in 2003]. http://www.xfml.org/spec/1.0.html.

VATANT, B. (2004) Published subject nocturne at XML 2004. 2004-11-09. Message to the mailing list : public-esw-thes@w3.org. http://esw.w3.org/mt/esw/archives/000094.html.

VELLUCCI, S. L. (1999) "Metadata and authority control", *Library Resources and Technical Services*, 44 (1) 1999, 33-43.

VELTMAN, K. H. (2001) "Syntactic and semantic interoperability : new approaches to knowledge and the semantic web", *The New Review of Information Networking*, 7 2001, 159-183.

VELTMAN, K. H. (2002) "Challenges for a Semantic Web", Proceedings of the International Workshop on the Semantic Web 2002 (at the Eleventh International World Wide Web Conference), Honolulu, Hawaii, May 7, 2002. http://semanticweb2002.aifb.uni-karlsruhe.de/proceedings/Position/veltmann.pdf.

VELTMAN, K. H. (2004) "Towards a Semantic Web for Culture", *Journal of Digital Information*, 4 (4) 2004, article no. 255. http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Veltman/.

VICKERY, B. C. (1959) "Notational symbols in classification : parts I-VI", *Journal of Documentation*, 8 (1) 1952, 14-31; 12 (2) 1956, 73-87; 13 (2) 1957, 72-77; 14 (1) 1958, 1-11; 15 (1) 1959, 12-16.

VICKERY, B. C. (1960) Faceted classification : a guide to the construction and use of special schemes, London : Aslib, 1960.

VICKERY, B. C. (1961) "The Universal Decimal Classification and technical information indexing", *UNESCO Bulletin for Libraries*, 15 (3) 1961, 126-138.

VICKERY, B. C. (1971) "Structure and function in retrieval languages", *Journal of Documentation*, 27 (2) 1971, 69-82. [Reprinted in: From classification to "knowledge organisation" : Dorking revisited or "past is prelude". Edited by A. Gilchrist. The Hague : FID, 1997. (FID 714), 23-34.]

VICKERY, B. C. (1972) "Classificatory principles in natural language indexing systems", *Classification in the 1970s*. [Edited] by Arthur Maltby. London : Clive Bingley, 1972, 119-141.

VICKERY, B. C. (1986) "Knowledge representation : a brief review", *Journal of Documentation*, 42 (2) 1986, 145-159. [Reprinted in: From classification to "knowledge organisation" : Dorking revisited or "past is prelude". Edited by A. Gilchrist. The Hague : FID, 1997. (FID 714), 35-48.]

VICKERY, B. C. (1990) "Classificatory principles in intelligent interfaces", *Tools for knowledge organization and the human interface : proceedings of the 1st International ISKO Conference, Darmstadt, 14-17 August 1990*. Edited by R. Fugmann. Frankfurt/Main : Indeks Verlag, 1990. (Advances in knowledge organization 1), 14-20.

VICKERY, B. C. (1997) "Issues in knowledge organization", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 180-182.

VICKERY, B. C (1997a) "Ontologies", *Journal of Information Science*, 23 (4) 1997, 277-286.

Visualizing subject access for 21st century information resources : [proceedings of 34th Annual Clinic on Library Applications of Data Processing, Champaign, 1997]. (1998) Edited by P. Atherton Cochrane, E. H. Johnson. Champaign, IL : Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 1998.

VIZINE-GOETZ, D. (1996) "Online classification : implications for classifying and document[-like object] retrieval", *Knowledge organization and change : proceedings of the Fourth International ISKO Conference, Washington, DC, 15-18 July 1996*. Edited by R. Green. Frankfurt/Main : Indeks Verlag, 1996. (Advances in knowledge organization 5). 378-382.

VIZINE-GOETZ, D. (1997) "From book classification to knowledge organization : improving Internet resource description and discovery", *Bulletin of the American Society for Information Science,* Oct/Nov'97. http://www.asis.org/Bulletin/Oct-97/vizine.htm.

VIZINE-GOETZ, D. (1998) "Dewey as an Internet subject guide", *Structures and relations in knowledge organization : proceedings of the Fifth International ISKO Conference, Lille, 25-29 August 1998*. Edited by W. Mustafa Elhadi, J. Maniez, S. Pollitt. Würzburg : Ergon Verlag, 1998. (Advances in knowledge organization 6), 191-197.

VIZINE-GOETZ, D. (1998a) "OCLC investigates using classification tools to organize Internet data", *Visualizing subject access for 21st century information resources : [proceedings of 34th Annual Clinic on Library Applications of Data Processing, Champaign, 1997]*. Edited by P. Atherton Cochrane, E. H. Johnson. Champaign, IL : Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 1998, 93-105.

VIZINE-GOETZ, D. (1999) "Using library classification schemes for Internet resources", *OCLC Internet Cataloging Project Colloquium*, 1999. http://www.oclc.org/oclc/colloq/v-g.htm.

VIZINE-GOETZ, D. (2003) Terminology services. Presentation at ECDL conference, 21 August, 2003. http://staff.oclc.org/~vizine/ECDL2003NKOS/ECDL03NKOS_vizine-goetz.ppt.

VIZINE-GOETZ, D. et al. (2004) "Vocabulary Mapping for Terminology Services", *Journal of Digital Information*, 4 (4) 2004, article no. 272 .

http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Vizine-Goetz/.

VIZINE-GOETZ, D. et al. (2004a) Terminology Services [presentation]. RAC meeting, September 2004.

Vocabulary ML (2000) : Metacode strawman DTD. NKOS, The Vocabulary Mark-up Language. NKOS, 2000. http://nkos.slis.kent.edu/.

WAGNER, G. K. (2000) "Why German libraries will never use the UDC", *Extensions & Corrections to the UDC*, 22 2000, 11. http://www.udcc.org/notes.htm#german.

WAJENBERG, A. S. (1983) "MARC Coding of DDC for Subject Retrieval", *Information Technology and Libraries,* 2 (3) 1983, 246-251.

WALLIS, J.; BURDEN, P. (1995) "Towards a classification-based approach to resource discovery on the web", *4th International W4G Workshop on Design and Electronic Publishing, Abingdon, England, 20-22 November 1995*. http://www.scit.wlv.ac.uk/wwlib/position.htm.

WARNER, J. (2000) "Can classification yield an evaluative principle for information retrieval?", *The future of classification*. Edited by R. Marcella, A. Maltby. Aldershot : Gower, 2000, 33-40.

WAUGH, A. (1998) "Specifying metadata standards for metadata tool configuration", *Proceedings of the seventh international conference on World Wide Web 7 table of contents, Brisbane, Australia, 1998*. Amsterdam : Elsevier, 1998, 23-32. http://www7.conf.au/programme/fullpapers/1913/com1913.htm.

WEIBEL, S. (1995) "Metadata : the foundations of resource description", *D-lib Magazine*, July 1995. http://www.cnri.reston.va.us/home/dlib/July95/07weibel.html.

WEIBEL, S. (1997) "The Dublin Core : a simple content description model for electronic Resources", *Bulletin of American Society for Information Science*, 24 (1)

1997, 9-11.

WEIBEL, S. (1997a) "Image description on the Internet : a summary of the CNI/OCLC Image Metadata Workshop : September 24-25, 1996, Dublin Ohio", *D-lib Magazine*, January 1997. http://www.dlib.org/dlib/january97/oclc/01weibel.html.

WEIBEL, S. et al. (1995) "OCLC/NCSA Metadata Workshop report". http://www.oclc.org:5046/oclc/research...ences/metadata/dublin_core_report.html.

WEIBEL, S.; IANNELLA, R.; CATHRO, W. (1997) "The 4th Dublin Core Metadata Workshop : report : DC-4, March 3-5, 1997, National Library of Australia, Canberra", *D-lib Magazine*, June 1997.
http://www.dlib.org/dlib/june97/metadata/06weibel.html.

WEINBERG, B. H. (1995) "Library classification and information retrieval thesauri : comparison and contrast", *Cataloging & Classification Quarterly*, 19 (3-4), 1995, 21-44.

WEINBERG, B. H. (1996) "Complexity in indexing systems : abandonment and failure : implications for organizing the Internet", *ASIS 1996 Annual Conference proceedings*. http://www.asis.org/annual-96/ElectronicProceedings/weinberg.htm.

WEINER, M. L.; LIDDY, E. L. (1995) "Intelligent text processing, and intelligence tradecraft", *The Journal of AGSI*, July, 1995. [Reprint : From classification to "knowledge organisation", Dorking revisited or "past is prelude". Edited by A. Gilchrist. The Hague : FID, 1997. (FID 714), 161-168.]

WEINHEIMER, J. L. (2000) "Metadata creation : down and dirty". Princeton University, May 2000.
http://www.princeton.edu/~jamesw/mdata/MetadataCreation.html.

WELLISCH, H. (1971) "Reorganization of the UDC", *Nachrichten für Dokumentation*, 22 (2), 1971, 55-63.

WELLISCH, H. (1976) "Uskladjivanje naucnih i tehnickih termina UDK i tezaurusa inzenjeringa i naucne tehnologije (TEST) [A concordance between UDC and TEST (Thesaurus of Engineering and Scientific terms : results of a pilot project]", *Proceedings of the international symposium "UDC in Relation to Other Indexing Languages" held in Herceg Novi, Yugoslavia, June 28-July 1, 1971*. Beograd *:* Jugoslovenski centar za tehnicku i naucnu dokumentaciju; The Hague : FID, 1976, 283-303.

WHEATLEY, A. (2000) "Subject trees on the Internet : a new role for bibliographic classification?", *Internet searching and indexing : the subject approach*. Edited by A. R. Thomas, J. R. Shearer. New York [etc.] : The Haworth Information Press, 2000, 115-141.

WIERENGA, K. (1996) "DESIRE : Development of a European Service for Information on Research and Education", *Ariadne*, 5 September, 1996. http://www.ariadne.ac.uk/issue5/desire/.

WILL, L. D. (1997) "Data structure and indexing for museum collections management", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 36-41.

WILLIAMSON, N. J. (1990) "The UDC : its future", *The UDC : essays for a new decade*. Edited by A. Gilchrist, D. Strachan. London : Aslib, 1990, 29-32.

WILLIAMSON, N. J. (1994) "The future revision of the UDC", *Extensions & Corrections to the UDC*, 16 1994, 19-27.

WILLIAMSON, N. J. (1996) "Deriving a thesaurus from restructured UDC", *Knowledge organization and change : proceedings of the Fourth International ISKO Conference, 15-18 July 1996, Washington DC*. Edited by R. Green. Frankfurt/Main : Indeks Verlag, 1996. (Advances in knowledge organization 5), 370-377.

WILLIAMSON, N. J. (1997) "Knowledge structures and the Internet", *Knowledge* *organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 23-27.

WIMMER, F. (2000) "The apostrophe in UDC : a critical examination of an unwelcome development", *Extensions & Corrections to the UDC,* 22 2000, 18-27.

WOLFRAM, D. (1996) "Inter-record linkage structure in a hypertext bibliographic retrieval system", *Journal of the American Society for Information Science*, 46 (10) 1996, 765-774.

WOLFRAM, D.; VOLZ, A.; DIMITROFF, A. (1996) "The effect of linkage structure on retrieval performance in a hypertext-based bibliographic retrieval system", *Information Processing and Management*, 32 (5) 1996, 529-541.

WOODS, E. W. (1996) "Requirements for a format for classification data : final report July 1996", *IFLA Sections on Classification and Indexing and Information Technology, 1996*. http://www.ifla.org/VII/s29/projects/appc.htm.

WORSFORD, E.; HOFFMAN, P.; HIOM, D. "Developing multilingual subject gateways : an issues paper written as part of the DESIRE Cataloguing Project". http://www.sosig.ac.uk/desire/lang/language.htm.

YI, H.; ZHANG, J. (1997) "The role of classification in the organization of information under the networking environment", *Knowledge organization for information retrieval : proceedings of the Sixth International Study Conference on Classification Research, London, 16-18 June 1997*. The Hague : FID, 1997. (FID 716), 183-185.

ZENG, M. L.; CHAN, L. M. (2004) "Trends and issues in establishing interoperability among knowledge organization systems", *Journal of the American Society for Information Science and Technology,* 55 (5) 2004, 377-395.

APPENDICES (1-10)

1. NETWORK KOS EXCHANGE STANDARDS

1.1 Web-based ontology languages

DARPA Agent Markup Language (DAML) is designed specifically to become a language and tool for facilitating the concept of the Semantic Web. DAML is a typical example of a standard that uses a Web-compatible language along with a reasoning paradigm developed in the field of AI.

DAML+OIL is a combination of a Web language, description logics, and a frame reasoning system as defined by The Ontology Inference Layer (OIL). It provides a rich set of constructs with which to create ontologies and to markup information so that it is machine-readable and understandable. DAML is compatible with the RDF Schema (RDFS), and includes precise semantics for describing term meanings (Fensel, 2000).

Ontology Web Language (OWL) is a semantic mark-up language developed by W3C. It is derived from DAML+OIL for the purpose of the Web ontology creation and exchange (http://www.w3.org/TR/owl-ref/). OWL contains three 'sub-languages' characterised by different levels of complexity: OWL Lite, OWL DL and OWL Full.

Simple HTML Ontology extension (SHOE) - which is an HTML-based knowledge representation language that offers categories, relationships, attributes, inferences, etc. that can be defined by ontologies. SHOE provides a relatively rich level of semantics and abilities, which enables Web designers to embed documents with information about the overall "content" of those documents. SHOE also allows agents to make automatic inferences about the data they learn, provides a hierarchical categorization scheme and a sophisticated ontology mechanism designed specifically for the needs of the Web (http://www.cs.umd.edu/projects/plus/SHOE/).

1.2 Standards for terminology exchange

Open Lexicon Interchange Format (OLIF) is an XML compliant, user-friendly, format for exchanging terminological and lexical data. The OLIF (version 2.0) lexicon and terminology exchange standard is currently under development within the OLIF Consortium, a collaborative group of industrial firms active in the field of language technology (http://www.olif.net/).

ISO 12200:1999 - MAchine-Readable Terminology Interchange Format (MARTIF) is an SGML-based format for data interchange among concept-oriented terminological databases. MARTIF is intended for interchange between partners (e.g., two translation companies) who know about each other and are able to "negotiate" details of the format to minimize information loss.

ISO 16642:2003 - Terminological Markup Framework (TMF) is built on OLIF and MARTIF, which are primarily focused on lexical data. TMF is, however, more network-orientated and includes features for conceptual and ontological aspects of terminology data (Romary, 2001). TMF defines underlying structures and mechanisms needed for computer representation of terminologies, regardless of any specific format. The purpose of this standard is to express constraints on the representation of computerised terminology and to maintain interoperability between representations. One specific representation format generated from TMF is Terminological Mark-up Language (http://www.loria.fr/projets/TMF/).

1.3 Standards for library KOS

MARC Authority Formats provide support to classification authority data. Their application is heavily limited by their bibliographic format encoding and, so far, an alternative Web-orientated encoding does not exist. They are tailored for library system applications with particular classification systems in mind and from the point of view of data modelling they are more suitable for database applications than for the Web environment. Representatives are *MARC 21 Concise Format for Classification Data* (http://www.loc.gov/marc/classification/eccdhome.html) and *Concise UNIMARC Format for Classification Data* (http://www.ifla.org/VI/3/p1996-1/concise.htm).

The Zthes profile - Zthes is a Z39.50 profile for thesaurus navigation. The profile describes an abstract model for representing and searching thesauri (e.g. hierarchies of terms as described in ISO 2788: 1986) and specifies how this model may be implemented using the Z39.50 protocol. It also suggests how the model may be implemented using other protocols and formats: a Zthes DTD for XML is provided as an appendix to the profile. Real Zthes datasets have been exchanged in the form of XML documents conforming to this DTD (Taylor M., 2001).

1.4 E-learning vocabulary standard

Vocabulary definition exchange (VDEX) is an IMS Global Learning Consortium specification that defines a grammar for the exchange of value lists of various classes (i.e. any collections of terms denoted as "vocabulary"). VDEX can be used for the exchange of simple machine-readable lists of values, or terms, together with information that may aid a human being in understanding the meaning or applicability of the various terms.

VDEX XML syntax can be used for strictly hierarchical schemes. The IMS Technical Board approved the VDEX Version 1 Final Specification in February 2004 (http://www.imsglobal.org/vdex/).

1.5 General (not domain specific) standards for vocabulary exchange

ISO/IEC 13250:2000 Topic Maps (XTM) is a specification that provides a model and grammar for representing the structure of information resources used to define topics and associations (relationships) between topics. Web-based Topic Maps, called XML Topic Maps (XTM) are developed in order to facilitate the use of the topic maps paradigm on the Web, and to help realise its potential for finding and managing information (http://www.topicmaps.org/xtm/1.0/). Topics have their characteristics represented within limited contexts in which they are given their name, resource and relationship characteristics. One or more interrelated documents employing this grammar are called a 'topic map'.

The Vocabulary Markup Language (Voc-ML). A NISO workshop on Electronic Thesauri: Planning for a Standard held in 1999 (Milstead, 1999) concluded (amongst other things) that there was a need for a metadata content standard for the description of knowledge organisation systems. NKOS has since then defined a set of attributes for the description of knowledge organisation systems, and developed a draft XML DTD known as the Vocabulary Markup Language (Vocabulary ML, 2000).

The schema includes Dublin Core metadata that would describe the knowledge organisation systems being encoded. It also defines tags and syntax for uniquely identifying each term, its relationship to other terms, and provides place for descriptive information like scope notes and definitions. It is hoped that the schema, when finalised, will work for a range of different types of system, including authority files, hierarchical thesauri, classification schemes, digital gazetteers and subject heading lists (Hodge, 2000).

eXchangeable Faceted Metadata Language - XFML Core (http://www.xfml.org/spec/1.0.html). The purpose of the XFML format is the use and exchange of a faceted vocabulary. In essence, similar to the XTM, XFML is "a model to express topics, organised in hierarchies or trees within mutually exclusive containers called facets" and enables vocabulary to be published in an XML format. It is possible to built connections between different XFML topic maps, by indicating that a topic in one map is equal to a topic in another map (Van Dijk, 2003).

SKOS – Simple Knowledge Organization System is a standard and specification (SKOS Core) for expressing knowledge organisation systems (KOS) in a machine understandable way. SKOS is a development by the W3C SWBP-WG Thesaurus Task Force within the SWAD-Europe project. SKOS Core is a model for expressing the basic structure and content of conceptual schemes. SKOS Core is a 'conceptual scheme' or 'concept scheme', defined here as: a set of concepts, optionally including statements about semantic relationships between those concepts.

SKOS uses a flexible XML/RDF syntax and is meant to be applied not only for thesauri but also for taxonomies, glossaries, Web directories etc. It is meant to be used as an ontology and is complementary to Ontology Web Language (OWL). The SKOS core provides a framework for publishing KOS terms and their relationships in order to support searching and browsing but is also supposed to support mapping and linking between different KOS. The SKOS Core was declared in 2004 to be an 'open' development (http://www.w3.org/2004/02/skos/).

BS 8723 - Structured vocabularies for information retrieval - guide. This is a new British Standard in development that will supersede the existing British and ISO standards for thesaurus establishment and development: ISO 2788-1986, ISO 5964-1985. The standard is planned to consist of five parts: (a) definitions; (b) guidance for creation of thesauri, electronic functions of thesauri and thesaurus management software; (c) guidance for creation of other types of structured vocabularies (classification schemes, search thesauri, subject heading lists, taxonomies and ontologies); (d) guidelines for interoperability between vocabularies, mapping etc. and (e) protocol and formats for vocabulary exchange (Dextre Clarke, Gilchrist & Will, 2004).

2. SUBJECT GATEWAYS USING UDC

2.1 NISS - Directory of Network Resources

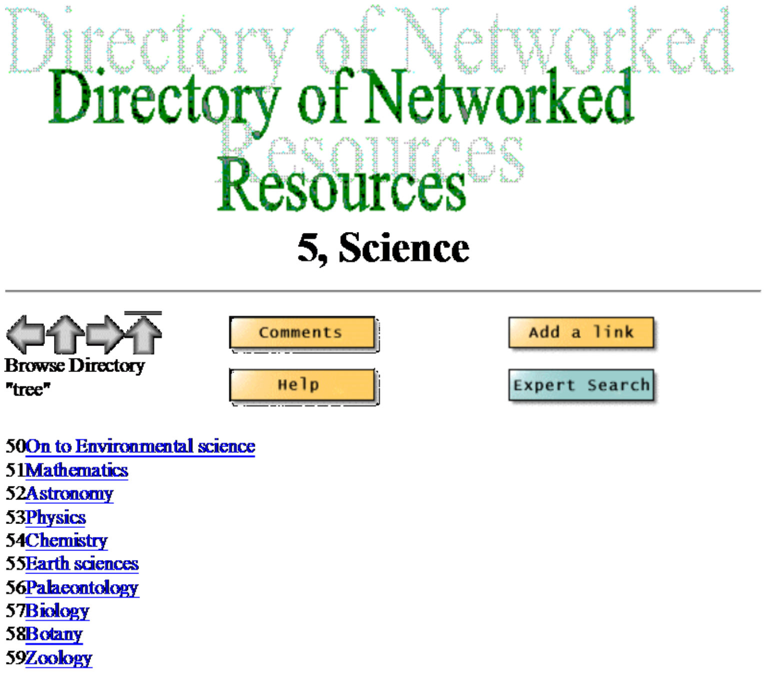


Figure 0.1 Interface 1999 (http://www.niss.ac.uk)

Figure 0.2 Interface 2004 (http://www.niss.ac.uk)

2.2 SOSIG - Social Sciences Gateway

Psychology
- Applied Psychology
  - Environmental Psychology
  - Health Psychology
- Cognition
  - Attention
  - Emotion and Motivation
  - Language, Speech, Linguistics
  - Memory and Learning
- Developmental Psychology
  - Child Psychology
  - Gerontology
  - Psychology of Adolescence
- Individual Differences
  - Intelligence
  - Personality
- Parapsychology
- Perception and Motor Functions
  - Visual Perception
- Physiological Psychology
- Psychological Disorders
  - Diagnosis and Treatment of Psychological Disorders

Figure 0.3 SOSIG Interface 1999 (http://www.sosig.ac.uk)

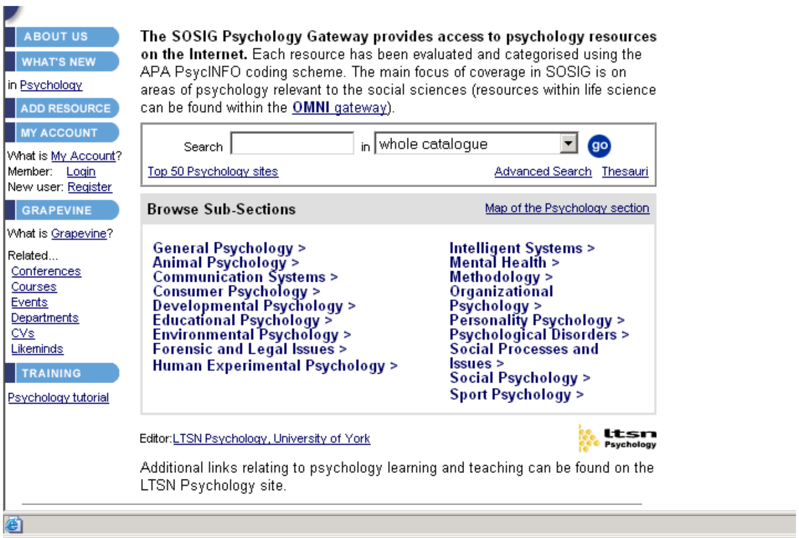


Figure 0.4 SOSIG interface 2004 (http://www.sosig.ac.uk)

2.3 Catalogue OKO

Figure 0.5 Slovenian catalogue of the Web resources (http://www.zrc-sazu.si/oko/)

2.4 Science Linkhouse

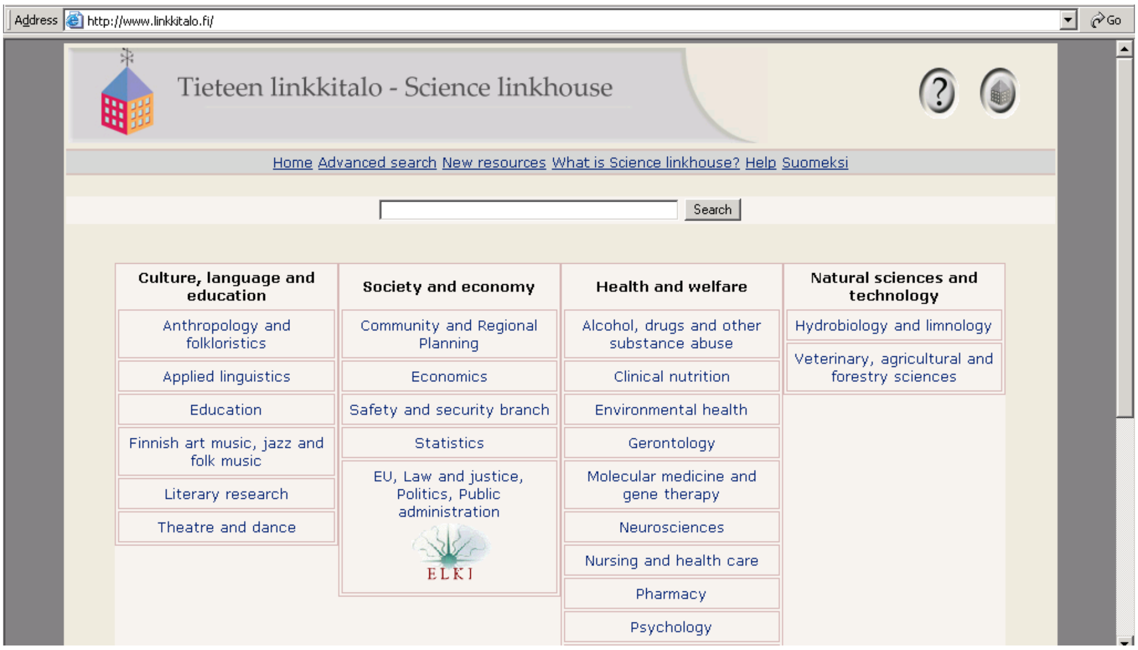


Figure 0.6 Science linkhouse, formerly FVL - Finnish Virtual Library (http://www.linkkitalo.fi/)

2.5 PORT -Maritime Information Gateway

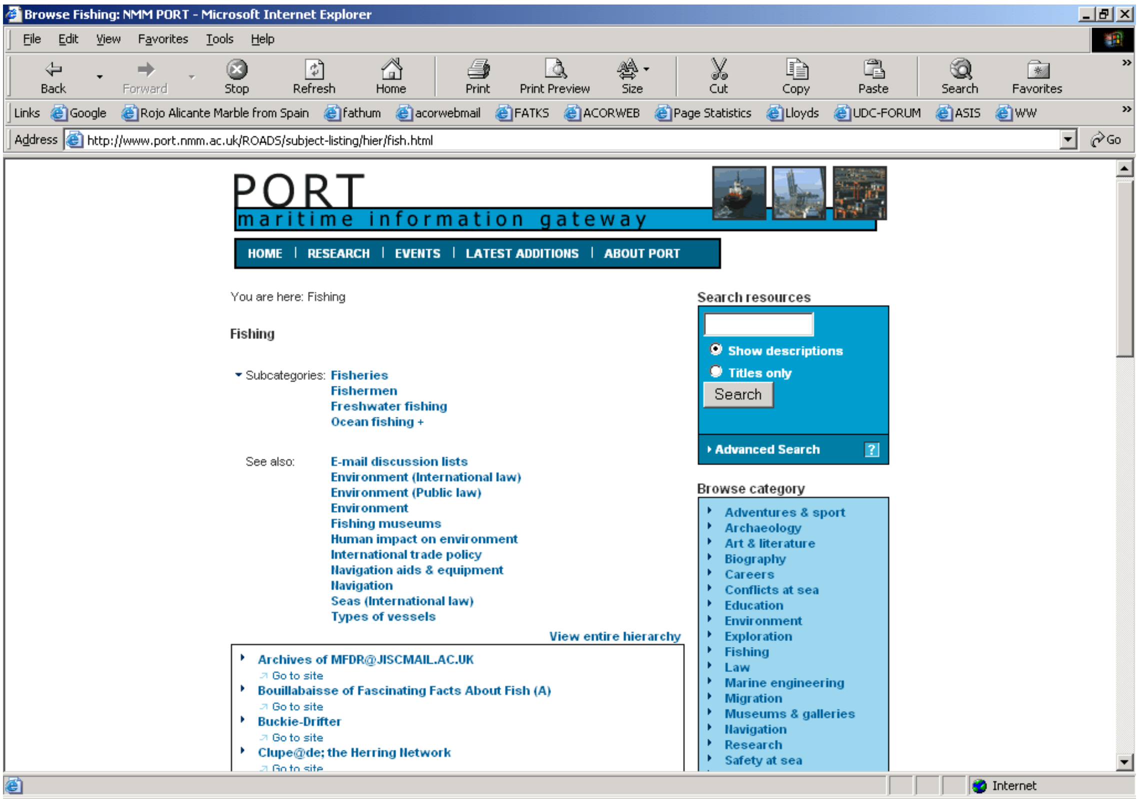


Figure 0.7: Browsing category 'fishing' on PORT (http://www.port.nmm.ac.uk/ROADS/subject-listing/hier/fish.html)

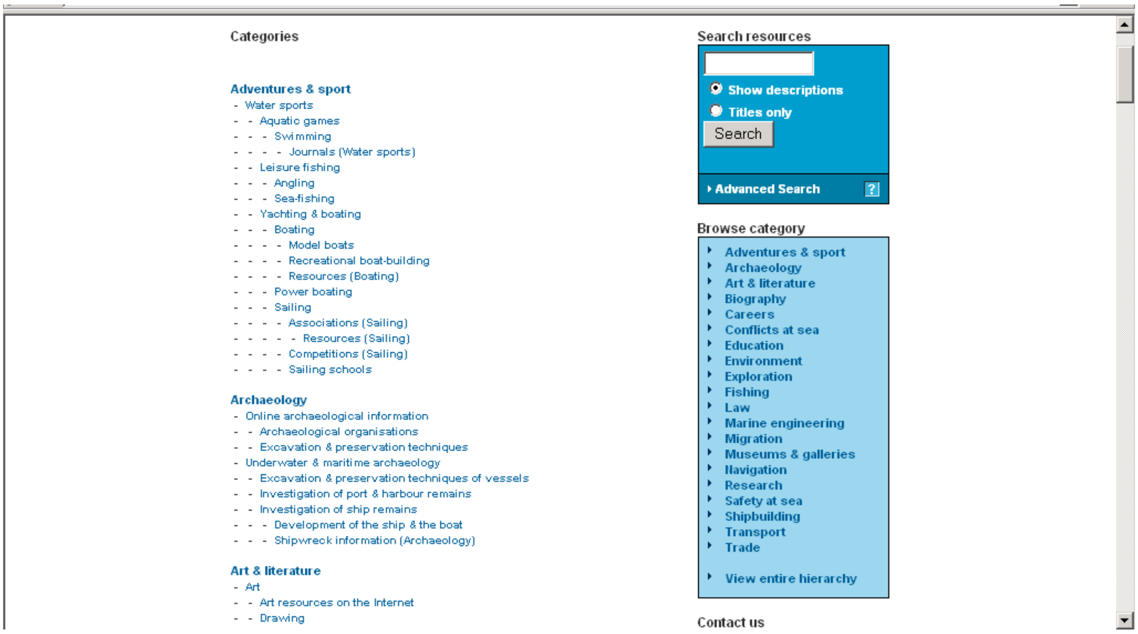


Figure 0.8: View of entire concept hierarchy on PORT

2.6 WAIS/World Wide Web service

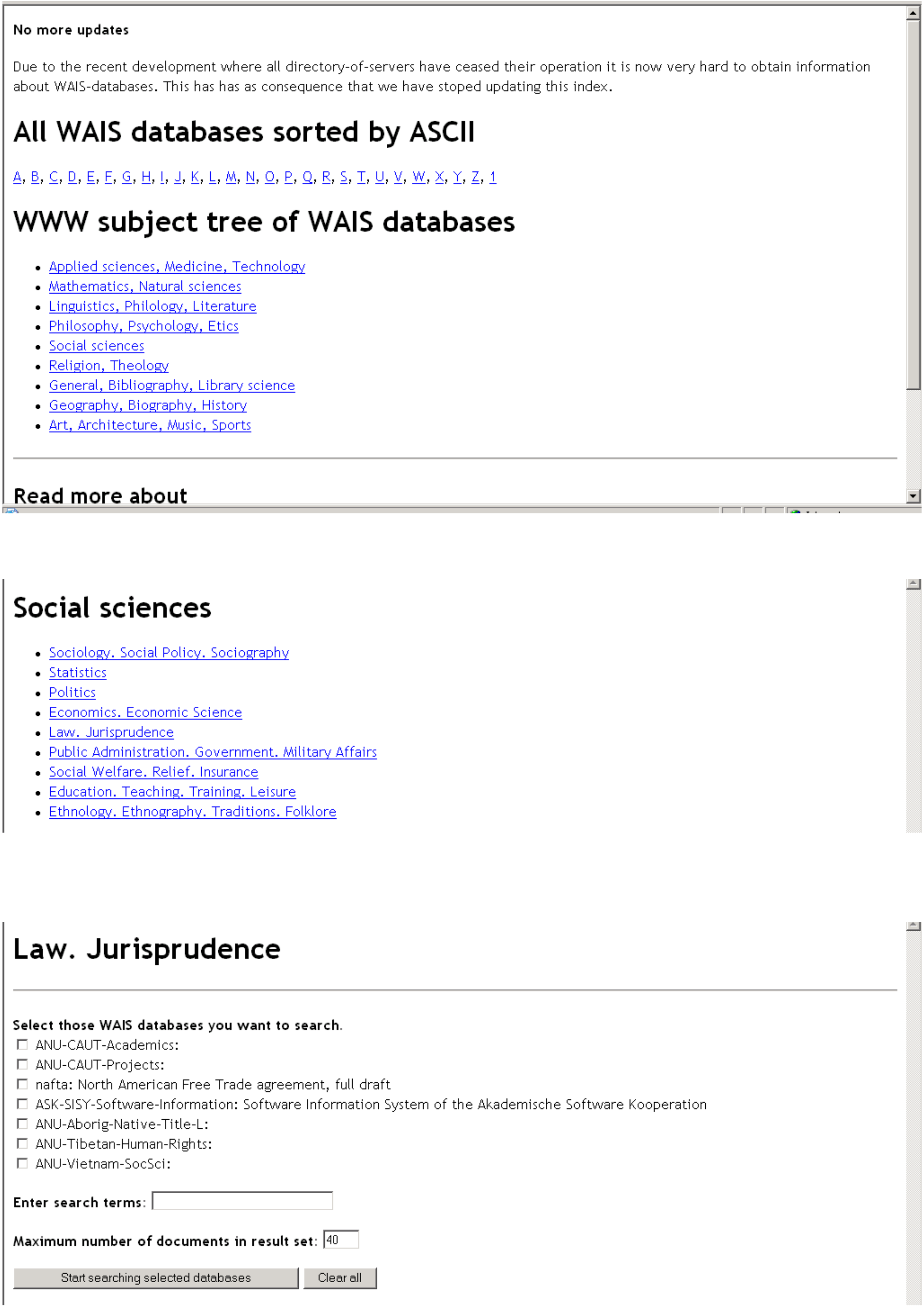
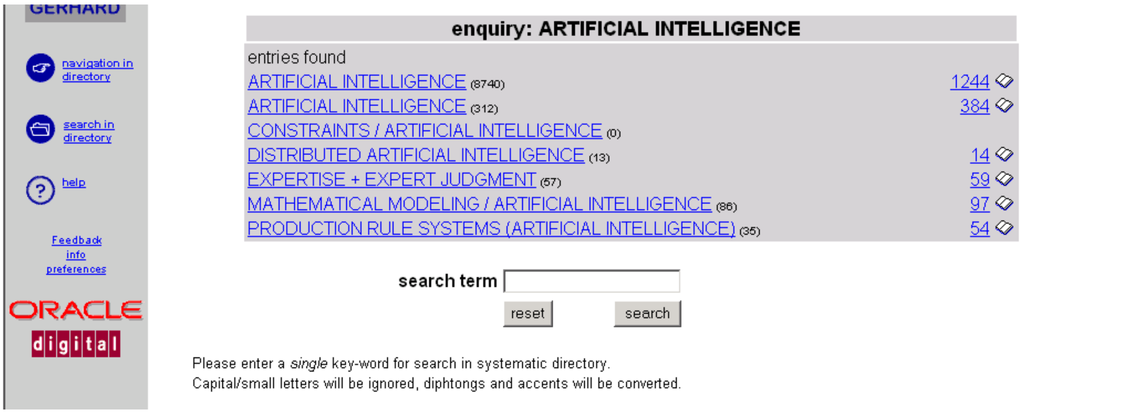
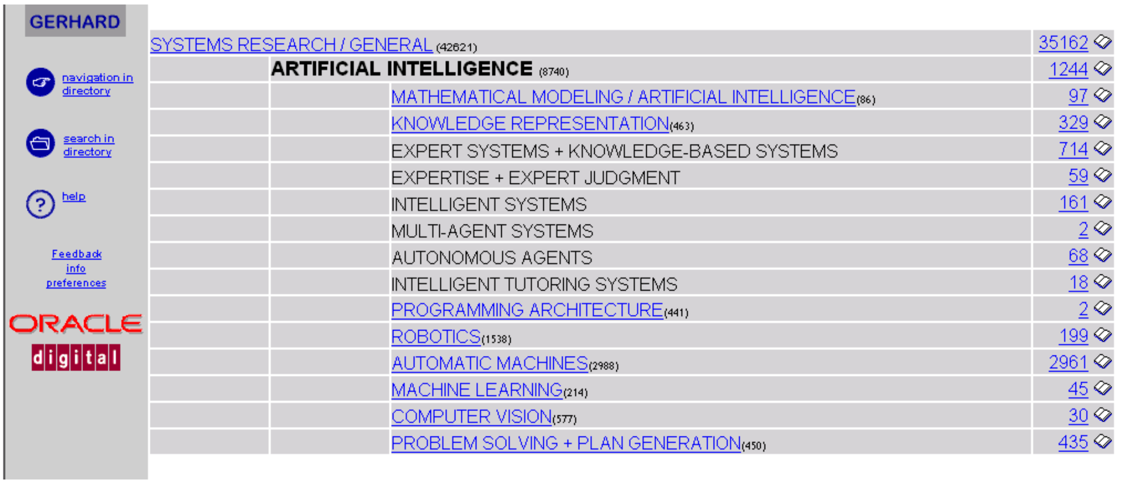


Figure 0.9 UDC browsing in WWW/WAIS (http://www.lub.lu.se/auto_new/UDC.html)

2.7 GERHARD





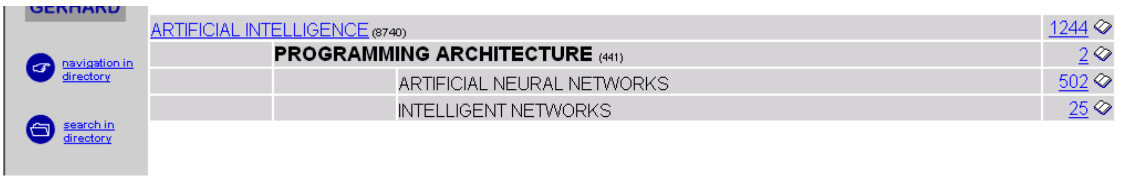


Figure 0.10 Three steps in directory browsing on GERHARD (http://www.gerhard.de)

3. UDC IN METADATA

3.1 Qualified Dublin Core Metadata Element Set (DCMES)

Metadata scheme	DCMES 1.1 and DCMES Qalifiers	
Encoding scheme	HTML, XML, XML/RDF	
Element group	Element / description	Scheme suggested
Content	Subject ID: Subject	
	Subject.Scheme	DC, UDC, LC, LCSH etc
	Coverage. Spatial	
	Coverage.Temporal	
	Type	DC type list:

Example 1 HTML embedded description

```
<html>
<heading>
...
<meta name="DC.Subject" content ="psychology">
<meta name="DC.Subject" content ="psychoanalysis">
<meta name="DC.Subject" scheme="UDC" content="159.96(73)">
<meta name="DC.Coverage.Spatial" content="United States">

</heading>
...
</html>
```

Example 2 standalone metadata XML

```
<?xml version = "1.0" encoding = "UTF-8"?>
<!--Created by Aida Slavic, 10/11/2002-->
<dc xmlns = "x-schema:ftp://ftp.myserver.co.uk/aida/dced_sbu.xdr">
...
<subject lang = "en-GB">psychology </subject>
<subject lang = "en-GB">psychoanalysis</subject>
<subject scheme="udc">159.96(73)</subject>
...
<coverage type="spatial">United States</coverage>

</xml>
```

Example 3 standalone metadata RDF

```
<?xml version="1.0" ?>
<RDF xmlns="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
     xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
     xmlns:dc="http://purl.org/dc/elements/1.1/"
     xmlns:dcterms="http://purl.org/dc/terms/">
  <Description>
    <dc:subject>
      <Description>
         <value>159.96</value>
         <rdfs:label>UDC</rdfs:label>
         <rdfs:isDefinedBy rdf:resource="http://www.udcc.org/"/>
      </Description>
    </dc:subject>
  </Description>
</RDF>
```

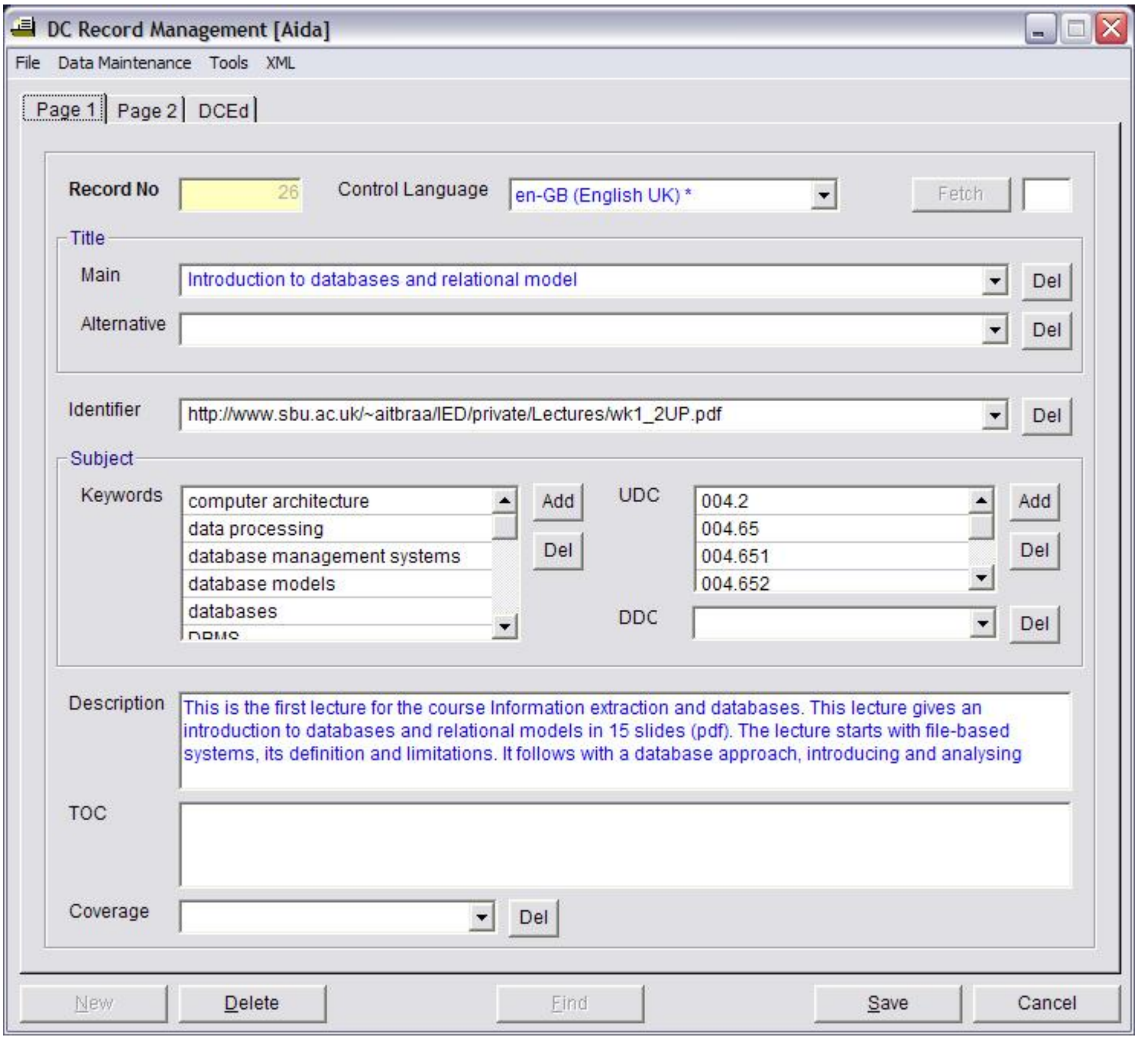


Figure 0.11 Use of UDC in Dublin Core metadata manager in EASEL project

3.2 IEEE/LOM - Learning Object Metadata

Metadata scheme	LOM	
Encoding language	XML	
Element group	Element / description	Scheme suggested
Subject data		
	1.4 Description	
1 General	1.5 Keywords	
	1.6 Coverage	TGN
9 External Classification	9.1 Purpose/ Discipline	
Example		

```
<classification>
<purpose>discipline</purpose>
<taxonpath>
<source>UDC</source>
<taxon>
<id>004</id>
<entry>Computer science and technology. Computing</entry>
</taxon>
<taxon>
<id>004.7</id>
<entry>Network</entry>
</taxon>
<taxon>
<id>004.73</id>
<entry>Network according to the area covered</entry>
</taxon>
<taxon>
<id>004.738</id>
<entry>Network interconnection</entry>
</taxon>
<taxon>
<id>004.738.5</id>
<entry>Internet</entry>
</taxon>
</taxonpath>
</classification>
```

3.3 AMICO Data Dictionary- Art Museum Image Consortium

Metadata scheme	Amico Data Dictionary
Encoding language	XML

Element group	Element / description	Scheme suggested
Classification Group	**Classification Scheme** Description: The classification scheme from which a term was chosen, followed by the unique identifier of the term within the scheme (if available). For the Art and Architecture Thesaurus, use AAT. For the Library of Congress Thesaurus of Graphic Materials use LCTGM:	AAT LCTGM
Object Material Group	**Materials and Techniques - Process / Technique - Term** Description: Single terms that index the processes and techniques used to create the work.	AAT Processes
	Materials and Techniques - Materials - Term Description: Single terms that index the materials used to create the work.	AAT Materials
Style Period Group	**Style Period Terms** Description: Index terms that characterize the style and/or period of the work of art.	AAT Style and Periods
Subject Matter Group	**Subject Matter - Pre - Iconographic Description** Description: A free text description of the generic subject of the work of art. Of-ness	AAT
	Subject Matter - Iconography Description: A free text description of the specific, named subject of the work of art. Aboutness	
	Subject Matter - Index Terms Description: Index terms that characterize the subject of the work of art	AAT, IconClass
Context Groups	**Context - Related Site / Place** Description: The names of any places that are contextually related to the work of art.	TGN
	Context - Time Period / Dates Description: The dates, times or periods of a particular context	

Example [from http://www.npaci.edu/DICE/AMICO/Demo/amico-objects.xml]

```
<am_object>
<AID>AIC_.1910.238</AID>
<OTY>Mummy Goods</OTY>
<CLG>
              <CLT>Mummy Goods</CLT>
</CLG>
     <CLG>
            <CLT>Ancient Art</CLT>
</CLG>
     <CLG>
       <CLT>European Decorative Arts and Sculpture and Ancient Art</CLT>
       </CLG>
      <OTG>
             <OTN>Mummy Case of Paankhenamun</OTN>
              <OTT>preferred</OTT>
       </OTG>
       <OPD />
       <OPA />
        <MET>H:  170.2  cm  (67  in.);  W.:  43.2  cm  (17  in.);  D.:  31.7  cm  (12-1/2
in.)</MET>
        <OMG>
       <OMD>Cartonnage (gum, linen, and papyrus), gold leaf, pigment</OMD>
</OMG>
</am_object>
```

3.4 EAD - Encoding Archive Description

Metadata scheme	EAD	
Encoding language	XML	
Element group	Element / description	Scheme suggested
EAD Header	<notestmt><subject>	
EAD Achdesc	<geogname> (spatial) <unitdate> (temporal) <controlaccess><subject>	

Example 1

```
<archdesc level="collection" langmaterial="eng">
<did>[...]</did>
<controlaccess>
<head>INDEX TERMS</head>
<note><p>These records are indexed under the following headings in the catalog
of the Minnesota Historical Society.  Researchers wishing to find related materials should
search the catalog under these index terms.</p></note>
<controlaccess>
<head>Organizations:</head>
<corpname>Board of Game and Fish Commissioners of Minnesota.</corpname>
</controlaccess>
<controlaccess>
<head>Topics:</head>
<subject>Fishery law and legislation--Minnesota.</subject>
<subject>Game-law--Minnesota.</subject>
<subject>Law enforcement--Minnesota.</subject>
</controlaccess>
</controlaccess>
[other possible elements and text... ]
</archdesc>
```

Example 2

```
<archdesc level="collection" langmaterial="eng">
<did>[...]</did>
<controlaccess><head>Selected Search Terms</head>
[other possible elements and text... ]
<controlaccess><head>Subjects:</head>
<subject encodinganalog="650">94"18"(410)</subject>
<subject encodinganalog="650">History of England in 19th century</subject>
</controlaccess>
</controlaccess>
[other possible elements and text... ]
</archdesc>
```

4. UDC IN OPACS

4.1 OPACs observation data

Table 9.1 Selection of Web OPACs providing subject access through UDC

COUNTRY	UDC USAGE[155]	LIBRARY	CAT. TYPE	URL	SYSTEM[156]		INTERFACE LANGUAGE	CODES[157]
					V/I	NAME		
AUSTRALIA	B	1. National Acoustic Laboratories Library	SINGLE	http://library.hearing.com.au/athcgi/athweb.pl	V	Athena	English	AU-NAL
AUSTRIA	C	2. Universitätsbibliothek der Technischen Universität Graz	UNION	http://tubss01.tu-graz.ac.at/ALEPH/DB1HUVVSJXKBD6V83RRK4KHA6IQRAEMUQVDKCXQ58PEVPYRMLG-00982/file/start-0	V	Aleph	German	AT-UT
AUSTRIA	C	3. Bibliothekskatalog der Universität Linz	UNION	http://aleph.edvz.uni-linz.ac.at/ALEPH/VUKXLJ2ND7R8RBKHH1XNFCGUKK14LFY82EDTISBT8MSKCR1MIR-00993/file/start-0	V	Aleph	German	AT-UL
BELGIUM	B	4. Universiteitsbibliotheek K.U.Leuven	UNION	http://bibserv.libis.kuleuven.ac.be/ipacbin/LibriVision/search_form.html?SESSION_ID=1100140389_7701&lv_action=LV_Search_Form&DB_ID=1&LDB_NAME=GV&HTML_SEARCH_TYPE=ADVANCED&L_SUBJECTS=0	V	DOBIS-LIBIS	Dutch, French	rBE-UL
CROATIA	A	5. Croatian National and University Library	UNION	http://www.nks.hr/	I	Crolist	Croatian	HR-NL
CROATIA	A	6. Zagreb City Library	UNION	http://www.kgz.hr/#	I	Zaki	Croatian	HR-KGZ
CROATIA	A	7. Croatian Academy of Sciences and Arts Library	SINGLE	http://161.53.55.45/web/start01.htm	I	ISIS	Cro., En.	HR-HA
CZECH R.	A	8. The National Library of the Czech Republic	UNION	http://sigma.nkp.cz/F/?func=file&file_name=find-b&local_base=skc&CON_LNG=ENG	V	Aleph	Czech, English	CZ-NL
ESTONIA	A	9. ESTER - Estonian Library Catalogues	UNION	http://helios.nlib.ee/search/	V	Innopac	Est.,Ru.,En.	EE-ES
FINLAND	B	10. Tampere University of Technology Library	UNION	http://tutcat.linneanet.fi/cgi-bin/Pwebrecon.cgi?LANGUAGE=English&DB=local&PAGE=First	V	Voyager	Finnish	FI-TUT
GERMANY	B	11. University Library of the Freiberg University of Mining and Technology	UNION	http://webopac.ub.tu-freiberg.de/	V	Libero	German	DE-FU

155 A - UDC is used in majority of libraries; B - UDC is used in certain type of libraries; C - UDC is very rarely used, in few libraries only
156 'V' indicates vendor library system, 'I' indicates in-house library system
157 This code is assigned to each Web OPAC for the purpose of this research and it will be used in research result presentation

COUNTRY	UDC USAGE[155]	LIBRARY	CAT. TYPE	URL	SYSTEM[156]		INTERFACE LANGUAGE	CODES[157]
					V/I	NAME		
HUNGARY	A	12. Miskolci Egyetem katalógusa	UNION	http://kvt20.lib.uni-miskolc.hu/webpac-bin/wgbroker.exe?2004111512344703502176+-access+top.Miskolcang	V	Horizon	Hun., En.	HU-MIS
IRELAND	B	13. Irish Management Institute	UNION	http://glas.imi.ie/GLASOPAC/index.asp	V	EOS -Glas		IE-IRI
ISRAEL	C	14. Technion Library	UNION	http://lib.technion.ac.il:4500/ALEPH/ENG/TEC/TEC/TEC/SCAN	V	Aleph	Heb., En.	IL-TEC
ITALY	B	15. Politecnico di Torino	UNION	http://opacbiblio.polito.it/ALEPH/-/start/libw	V	Aleph	It.	IT-POL
ITALY	B	16. Università di Genova	UNION	http://catalogo.sbi.genova.it/ALEPH/	V	Aleph	It.	IT-UG
LATVIA	A	17. Latvian National Library	UNION	http://195.13.129.13:8080/ALEPH/ESBYKPEE3U8DT8QYQDFMY2QDGIRLM6X2UGAYSFC5CVUJEF2RYU-00224/file/start-0	V	Aleph	Latv., En.	V-LNL
LITHUANIA	A	18. Lithuanian National Library	SINGLE	http://www.libis.lt:8080/en/welcome.html	V	Dobis Libis	Lith., En.	LT-LNL
NORWAY	B	19. Bibsys Library Data Centre	UNION	http://wgate.bibsys.no/search/pub?lang=N	I	Bibsys	No., En.	NO-BIB
PORTUGAL	A	20. National Bibliographic Database - PORBASE	UNION	http://sirius.bn.pt/sirius/sirius.exe	V	Sirius	Pt.	PT-POR
PORTUGAL	A	21. Fundação Calouste Gulbenkian, Biblioteca de Arte	SINGLE	http://www1.gulbenkian.pt/ipac-cgi/ipac.exe	V	Horizon	Pt.	PT-GUL
SLOVENIA; MACEDONIA, SERBIA & M.N.	A,A,A	22. Slovenian Union Cat.; Macedonian Union Cat.; Serbian Union Cat.; Monte Negro Union Cat.	UNION	http://www.cobiss.si; http://www.nubsk.edu.mk/cobiss/; http://vbs.nbs.bg.ac.yu/cobiss/; http://cnbct.cnb.cg.ac.yu/cobiss/	I	"COBISS"	Slov., Mac., Serb.	COBIS
SPAIN	A	23. Catàleg Col·lectiu de les Universitats de Catalunya	UNION	http://www.cbuc.es/ccuc/	V	VTLS	Sp.; Cat.	ES-UNI
SPAIN	A	24. Biblioteca de la Universidad Carlos III de Madrid	UNION	http://biblioteca.uc3m.es/uhtbin/cgisirsi/8t4cB3wD8c/90520072/61/1592/X	V	Unicorn	Sp.	ES-BUC
SWITZERLAND	B	25. Bibliothèque nationale suisse	UNION	http://www.coris.ch/newbns/Francais/saisie.asp	V	VTLS	Fr., Ger.	CH-BNS
SWITZERLAND	B	26. Eidgenössische Technische Hochschule (ETH)	UNION	http://opac.nebis.ch/ALEPH/39TKJTMJAYHRCS7G2139VII7LK731PB637XS13LN8DKAHUICJG-13604/file/start-ids	V	Aleph	Fr.; Ger., En.	CH-ETH
U.K	B	27. RIBA British Architectural Library	SINGLE	http://195.171.22.30/uhtbin/cgisirsi.exe/W2etq4E4dq/0/49	V	Unicorn	En.	UK-RIB
U.K.	B	28. School of Pharmacy Library	SINGLE	http://unicorn.ulsop.ac.uk/uhtbin/cgisirsi/Sat+Jun+26+16:26:32+BST+2004/0/49	V	Unicorn	En.	UK-SPL
U.K.	B	29. University of Surrey Library	UNION	http://opac.lib.surrey.ac.uk/www-bin/www_talis32	V	Talis	En.	UK-US
U.K.	B	30. University of Bath Library	UNION	http://www.bath.ac.uk/library/webcat/	V	Unicorn	En.	UK-UB

Table 1.2 Presence of IR functions in Web OPACs[158]

	IN-HOUSE SYSTEMS					VENDOR SYSTEMS																							
LIBRARY SYSTEM	bibsys	Crolist	COBSS	ISIS	Zaki	Aleph								Athena	EOS	Horizon		Innopac	Libero	Dobis Libis		Sirius	Talis	nicorn Sirsi				Voyager	VTLS	
OPAC CODES	NO-BIB	HR-NL	COBIS	HR-HA	HR-KGZ	AT-UT	AT-UL	CH-ETH	CZ-NL	IL-TEC	IT-POL	LV-LNL	IT-UG	AU-NAL	IE-IRI	HU-MIS	PT-GUL	EE-EST	DE-FU	BE-UL	LT-LNL	PT-POR	UK-US	ES-BUC	UK-RIB	UK-SPL	UK-UB	FI-TUT	CH-BNS	ES-UNI
1. ACCESS TO UDC[159]			• Si		• Si	• Si	• Si	• Bi	• Bi	• Bi	• Si	• Bi	• Bi	• Bc	• Bc	• Si	• Si	• Si	• Si	• Si	• Si	• Si	• Bi	• Si		• Bc	• Bc	• Bi	• Si	• Sc
2. EXPERT SEARCH [160]			•					•																						
3. UDC IN CALL NUMBER					•									•	•				•				•	•	•	•	•	•		•
4. UDC AS AN INDEX TERM	•	•	•	•	•	•	•	•	•	•	•	•	•			•	•	•	•	•	•	•						•	•	
5. LIST TO SELECT SEARCH TERMS	•					•	•	•								•	•											•	•	
6.UDC USED ALONGSIDE OTHER CLASSIFICATION	•												•							•										
7. UDC USED ALONGSIDE SUBJECT HEADINGS	•			•	•	•	•	•	•	•				•		•	•	•		•	•			•						

[158] Columns showing catalogues using UDC in call number only are shaded in grey
[159] Access to UDC: separate screen A-advanced search; Si - Search Index; Bi - Browse Index; Bc - Browse call number; Sc - Search call number
[160] Search using field codes and search syntax

	IN-HOUSE SYSTEMS					VENDOR SYSTEMS																								
LIBRARY SYSTEM	bibsys	Crolist	COBSS	ISIS	Zaki	Aleph								Athena	EOS	Horizon		Innopac	Libero	Dobis Libis		Sirius	Talis	nicorn Sirsi				Voyager	VTLS	
OPAC CODES	NO-BIB	HR-NL	COBIS	HR-HA	HR-KGZ	AT-UT	AT-UL	CH-ETH	CZ-NL	IL-TEC	IT-POL	LV-LNL	IT-UG	AU-NAL	IE-IRI	HU-MIS	PT-GUL	EE-EST	DE-FU	BE-UL	LT-LNL	PT-POR	UK-US	ES-BUC	UK-RIB	UK-SPL	UK-UB	FI-TUT	CH-BNS	ES-UNI
8. SUBJECT DIRECTORY BROWSE				•				•																				•	•	
9. SEARCH TO BROWSE	•					•	•	•	•		•	•	•	•	•		•	•	•	•			•	• cn		• cn	• cn	•	•	•
10. FREE BROWSE UDC INDEX	•			•		•	•	•	•	•	•		•	• cn	• cn			•								•	• cn	• cn	•	•
11. UDC HYPERLINK IN BIB. RECORD		•			•	•	ns	•	•	•	•	ns	•					ns		•	•		•	•	•	•				
12. LAUNCH SEARCH FROM AUTHORITY FILE								•						•					•	•									•	
13. BT/NT BROWSE UDC IN SEARCH RESULTS								•											•	•										
14. DEFAULT TRUNCATION	•					•	•		•	•		•	•	•	•	•		•	•			•	•	•		•	•	•	•	•
15. APPROXIMATE MATCH	•					•	•	•		•		•	•		•		•	•	•				•	•		•				
16. BOOLEAN SEARCH OF UDC INDEX[161]			s •		s •																								•	

[161] When Boolean search allows search of two independent UDC numbers (and not Boolean between parts of the same number) this is marked with "s"

	IN-HOUSE SYSTEMS					VENDOR SYSTEMS																								
LIBRARY SYSTEM	bibsys	Crolist	COBSS	ISIS	Zaki	Aleph								Athena	EOS	Horizon		Innopac	Libero	Dobis Libis		Sirius	Talis	nicorn Sirsi				Voyager	VTLS	
OPAC CODES	NO-BIB	HR-NL	COBIS	HR-HA	HR-KGZ	AT-UT	AT-UL	CH-ETH	CZ-NL	IL-TEC	IT-POL	LV-LNL	IT-UG	AU-NAL	IE-IRI	HU-MIS	PT-GUL	EE-EST	DE-FU	BE-UL	LT-LNL	PT-POR	UK-US	ES-BUC	UK-RIB	UK-SPL	UK-UB	FI-TUT	CH-BNS	ES-UNI
17.SEARCHING PARTS OF UDC[162]	•							•	•		•	*	*			•	*			•	•	*						*		
18. SEARCHING UDC CAPTION	•			•	•	•	•	•	•			•				•		•		•										
19. CLASSIFIED RESULT DISPLAY	•						•		•			•	•					•	•	•					•			•	•	
20. SHELFLIST ITEM DISPLAY							•	•						•	•								•	•		•	•			•
21. CORRECT SORT OF UDC									•										ns		ns	ns	•							
22. UDC CAPTION DISPLAY				•	•			•																					•	
23. HIERARCHICAL INDENTATION OF HITS																				•									•	
TOTAL NO. OF FUNCTIONS	**12** 53%	**2** 7%	**4** 17 %	**6** 26%	**8** 35%	**10** 43%	**11** 48%	**16** 69%	**11** 48%	**7** 30%	**6** 26%	**7** 30%	**9** 39%	**8** 35%	**7** 30%	**7** 30%	**6** 26%	**9** 39%	**8** 35%	**12** 53%	**5** 22%	**3** 13%	**8** 35%	**8** 35%	**3** 13%	**8** 35%	**6** 26%	**9** 39%	**12** 53%	**6** 26%

[162] In order to perform search on parts of complex numbers main auxiliaries are entered in separate field. This situation is marked as "*".

4.2 Interface examples

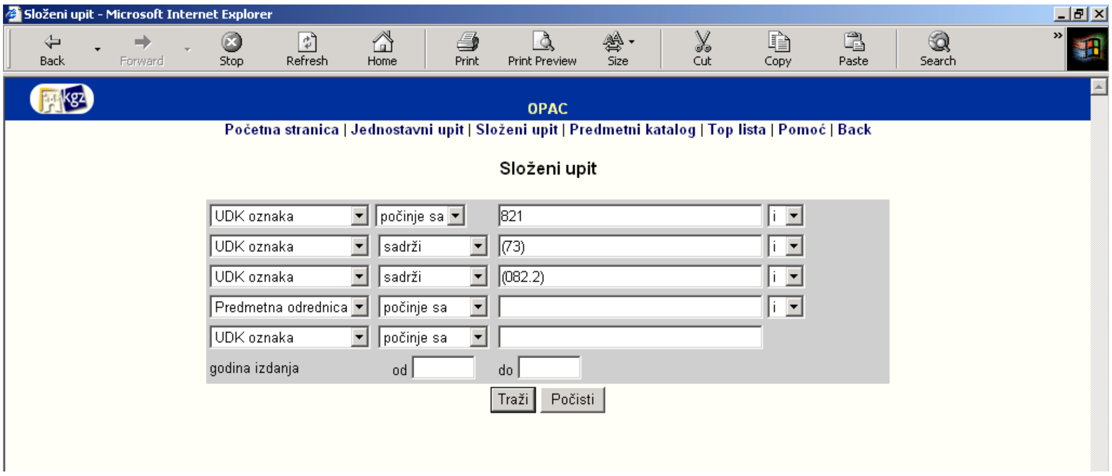


Figure 1.1 City library in Zagreb, Boolean search on parts of UDC[163]

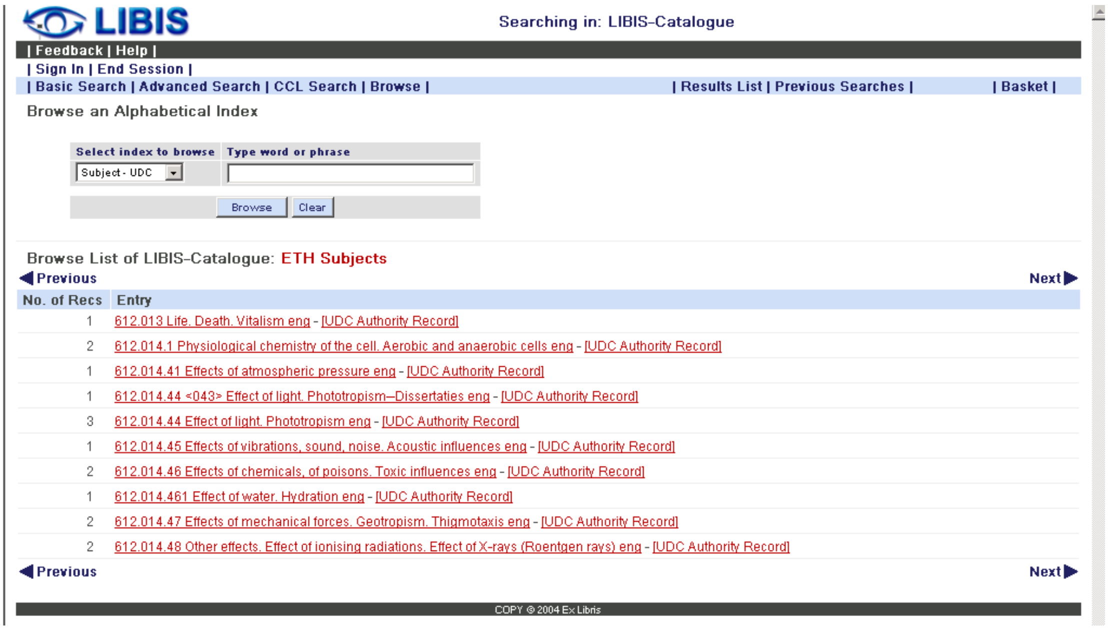


Figure 1.2 Browsing UDC in new catalogue of University of Leuven library[164]

---

[163] The screen capture shows a Boolean search of UDC numbers: *begins with* 821 AND *contains* (73) AND *contains* (082.2). This interface allows a search combination of UDC numbers, for each of which it is possible to determine a type of truncation (begins with, contains, ends with or exact match).

[164] The OPAC included in the observation was the Dobis Libis system. This library is now migrating to Aleph and is developing an authority file very similar to the ETH library in Zurich.

5. MARC CLASSIFICATION FORMATS

Table 1.3 MARC 21 Concise Format for Classification Data - outline

	MARC 21 Classification Format blocks
1xx Classification number and terms	
	153 Classification number
	Subfield codes
	Number portion
	≠a Classification number - single number or beginning of a span
	≠c Classification number - ending number of span
	≠z Table identification
	Caption portion
	≠h Caption hierarchy
	≠j Caption
	≠k Summary number span caption hierarchy
	Tracing and control subfields
	≠i Reference instruction phase
	≠t Topic
	≠w Control subfield
	154 - General explanatory index term (Information for Translators and Other Users)
	253/ 353/ 453/ 553Tracing and Reference Fields
	253 - Complex see reference
	353 - Complex see also reference
	453 - Invalid number tracing
	553 - Valid number tracing
	6XXNotes
	680 - Scope note
	681 - Classification example tracing note
	683 - Application Instruction Note
	684 - Auxiliary Instruction Note
	685 - History note
	70x - 75Index term fields
	700 - Index term--personal name
	710 - Index term--corporate name
	711 - Index term--meeting name
	720 - Index term--uncontrolled name
	730 - Index term--uniform title
	748 - Index term--chronological
	750 - Index term--topical
	751 - Index term--geographic name
	753 - Index term--uncontrolled
	754 - Index term--faceted topical terms
	76xNumber Building Fields
	761 - Add or divide like instructions
	762 - Table identification
	763 - Internal subarrangement or add table entry
	764 - Rule identification
	765 - Synthesized number components
	768 - Citation and preference order instructions
	8xx
	856 - Electronic location and access
	880 - Alternate graphic representation

Table 1.4 Concise UNIMARC Classification Format - schema outline

UNIMARC Classification Format
$4 Classification syntax and attribute codes **/0** Type of Relationship between class numbers **a** + **b** : **c** :: **d** [ ] **/1** Type of special auxiliary **a** .0 **c** - **d** ` **/2** Alphabetical extension **/3** Use of Auxiliary Table Numbers as Main Numbers
2--Heading block **250** Class number **$c** Ending Number of Span **$h** Caption Hierarchy **$j** Caption **$k** Summary Number Span Caption Hierarchy. **$z** Table Identification
3--Notes block **330** General Scope Note
4-- See Reference Block **453** Invalid Number Tracing
5--See Also Reference Block **553** Valid Number Tracing
66-Number Building Block **661-668** Number Building Fields **661** Add or Divide Like Instructions **662** Table Identification **663** Internal Subarrangement or Add Table Entry (Repeatable) **665** Synthesised Number Components **668** Citation and Precedence Order Instruction
70-- 75- Index Term Block **700-754** Index Terms **700** Index Term, Personal Name **710** Index Term, Corporate Name **720** Index Term, Family Name **730** Index Term, Uniform Title **750** Index Term, Topical **751** Index Term, Geographic Name **753** Uncontrolled Terms **754** Index Term - Faceted Topical Terms
8--Source Information Block **801** Originating Source **804** Authentication Code **820** Classification Example Tracing Note **825** Application Instruction Note **830** Auxiliary Instruction Note **856** Electronic Location and Access **875** History Note
9xxNational Use Block

Table 1.5 Comparison between UNIMARC and MARC 21 classification formats

UNIMARC (main heading field 250)		MARC 21 (main heading field 153)	
$a	Number Where Instructions are Found (Repeatable)	$a	Number where instructions are found--single number or beginning number of span (R)
$b	Base NumberContains the base class number to which one or more other numbers are added. In many instances it may be the same as subfield (Repeatable)	$b	Base number (R)The base classification number to which one or more other numbers are added. In many instances it may be the same as subfield $a. When two or more additions are used to build one number, the base number for the second and following 765 fields is the number resulting from the immediately preceding addition.
$a	When two or more additions are used to build one number, the base number for the second and following 665 fields is the number resulting from the immediately preceding addition. Repeatable.	$a	When two or more additions are used to build one number, the base number for the second and following 765 fields is the number resulting from the immediately preceding addition.
$c	Number Where Instructions are Found--Ending Number of Span (Repeatable)	$c	Number where instructions are found--Ending number of span (R)
$f	Facet Designator Contains extra character(s) to be added to class numbers that indicate facets. Repeatable.	$f	Facet designator (R) Extra character(s) to be added to classification numbers that indicate facets.
$r	Root Number Contains the initial digits of the pattern number or span when these digits are not added. If this subfield is present, subfield $s or $t must also be present to specify the digits that are added. Repeatable.	$r	Root number (R) The initial digits of the pattern number or span when these digits are not added. If this subfield is present, subfield $s or $t must also be present to specify the digits that are added.
$s	Symbols for Concepts Added from Class Number in Schedule or External Table These symbols may consist of alphanumerics and/or other symbols. Repeatable.	$s	Digits added from classification number in schedule or external table (R)
$t	Symbols for Concepts Added from Internal Subarrangement or Add Table These symbols may consist of alphanumerics and/or other symbols. Repeatable.	$t	Digits added from internal subarrangement or add table (R)
$u	Number Being Analysed Identifies the number being analysed. Subfield $u is repeated when the same 665 field is applicable to more than one number appearing in the same record. If the number being analysed is in field **153**[165] (Classification Number), indicated by value 0 in the first indicator, this subfield need not be used unless both numbers in a class number span are being analysed. Repeatable.	$u	Number being analyzed (R) Repeated when the same 765 field is applicable to more than one number appearing in the same record. If the number being analyzed is in field **153** (Classification Number), indicated by value 0 in the first indicator, this subfield need not be used unless both numbers in a classification number span are being analyzed.
$v	Number in Internal Sub-arrangement or Add Table Where Instructions are Found (Repeatable)	$v	Number in internal subarrangement or add table where instructions are found (R)
$w	Table Identification--Internal Subarrangement or Add Table (Repeatable)	$w	Table identification--internal subarrangement or add table (R)
$zTable Identification (Repeatable)		$zTable identification (R)	

[165] Field tag 153 should be 250 as in UNIMARC. This tag comes from MARC 21, from which the schema is obviously copied.

6. UDC MANAGEMENT

6.1 MRF statistics

Table 1.6 UDC MRF 2003 data population statistics

CLASSIFICATION FIELDS		
MRF FIELD	**NUMBER OF RECORDS**	**COMMENTS**
001 UDC number	66,733	This field can be considered as a unique primary key in a database. Field cannot be empty, or repeatable. Field content cannot be changed/altered
002 Table (no subfields) **Coded field value:** aTable 1a – Coordination. Extension b Table 1b – Relation. Subgrouping. Order-fixing c Table 1c – Com. aux. of language d Table 1d – Com. aux. of form e Table 1e – Com. aux. of place f Table 1f – Com. aux. of ethnic grouping g Table 1g – Com. aux. of time h Table 1h –notations from non-UDC source i Table 1i – Com. aux. of point-of -view [!] k Table 1k – Com. aux. of general charact. l Section II. Special auxiliary subdivisions M Main table	66,733 records contain data in this field Main numbers54,571 (17.493 of these are combination with special auxiliaries) Common auxiliaries12,162 Table 1a 1 Table 1b 1 Table 1c /language1,365 Table 1d/ form 359 Table 1e/place8,559 Table 1f/ethnic grouping 38 Table 1g/time 283 Table 1h/non-UDC source 1 Table 1i/point-of -view 0 Table 1k/general charact.1,553 Special auxiliary table 1	Main numbers can be simple, compound (main auxiliary with special auxiliary) and complex (main auxiliary with common auxiliary). Complex combination with colon is not allowed in the field 001. This type of main number combination can only be retrieved if field 004 (see below) is implemented. Common auxiliaries tables are coded. Table k which has the same code for the 4 auxiliary tables (properties, materials, processes, persons) Special auxiliary always appear attached to the main number, which can be considered as a 'base number'. Main number combined with special auxiliary will contain 'M' in the field 002 and will have code value in the field 003
003 Special aux. type **Coded field value:** A hyphen (–) auxiliary B point-nought (.0) auxiliary C single apostrophe (\`) auxiliary D other, e.g. final digits (…1/…9)	17,493 special auxiliaries (i.e. main numbers combined with special auxiliaries) - (hyphen) auxiliary5,141 .0 (point-nought) auxuliary11,814 \` (single apostrophe) auxiliary 574 others 23	3 records contain D in the field by mistake.
004 Combination type (no subfields)	[not implemented]	
010 Parallel derivation	902	Fields contain notation that denote hierarchy from which this number is derived. This field is applicable for numbers that are already devised based on parallel division
011 Instruction for parallel division Subfields ^a Source notation for the parallel division ^b Target notation (if not identical with the number in field 001) ^t English annotation ^g German annotation	924 source924 target109 English text34 German and English text2	This field is relevant for numbers that can be further built based on parallel division and contain instruction for that Tag ^a is missing from one source. Some field contains text and number, some contain only number
012 Special auxiliary supplied by parallel division Coded field value: A hyphen (–) auxiliary B point-nought (.0) auxiliary C inverted comma (\`) auxiliary D other	19 A2 B15 C3 D0	If parallel division should be used to create special auxiliaries then this field indicates the type of spec. aux. to be created
013 Special auxiliary supplied by annotation	118 records	If application note (v111) or scope note (v110) gives information

CLASSIFICATION FIELDS		
MRF FIELD	**NUMBER OF RECORDS**	**COMMENTS**
Coded field value A hyphen (–) auxiliary B point-nought (.0) auxiliary C single apostrophe (`) auxiliary D other, e.g. final digits (…1/…9)	- (hyphen) auxiliary35 .0 (point-naught) auxiliary76 ` (single apostrophe) auxiliary11 others/final digit 2	about application of special auxiliaries. This field denotes the type of special auxiliary in question.
100 Description Subfields ^eEnglish text ^gGerman text	66,700 English text66,700 German and English text2,233	Field of description is mandatory and is not repeatable which means that there should be 66,733 records containing data in this field. There are 33 records that miss this description in MRF 2003 - this has been reported and action has been taken.
105 Verbal examples **Subfields** ^eEnglish text ^gGerman text	8,229 English text8,225 German and Engish text 247	This field is simple extension of the field 100 Description. All records that contain this field should have text in English. There are three records that have only German text and one record that contains English text but is missing the tag. This was reported and should be corrected in the next edition of the MRF. Action taken.
110 Scope note **Subfields** ^eEnglish text ^gGerman text	775 English text774 German & English text 32	There are a number of scope notes that contain instructions that should go in the field 111. This was reported to revision in 2003 and 2004. No action yet taken. One record is missing tag ^e in front of English text.
111 Application note **Subfields** ^eEnglish text ^gGerman text	1,316 English text1,316 German and English text 59	There are number of instructions that are actually scope notes and should be in the field 110. This was reported to revision in 2003 and 2004. No action yet taken.
115 Combination examples **Subfields:** ^aDirect addition ^bColon (:) combination ^cFull notation Used if the first component differs from the notation in field 001. ^rReference(s) to UDC-number ^dEnglish description ^gGerman description ^nEnglish annotation ^yGerman annotation	2341 direct addition1,023 colon combination 935 full notation 384 reference 24 English description2,340 German and English description 91 English annotation27 German and English annotation1	One record has an incorrect tag in front of the English description. Four records use ^n by mistake. At least these mistakes have been reported and action taken.
120 Parallel division examples **Subfields** ^aNotation ^dEnglish description ^gGerman description	9 notations9 English description9 German and English description0	
125 References **Subfields** ^aNotation ^tEnglish annotation ^gGerman annotation	6,892 notations6,889 English text 142 German and English text 2	In general, a description is not given next to the references. This can be automatically displayed using description from the existing records. In 2004 action was taken to remove existing annotation that are actually descriptions - and to automate the process. Three records are missing the tag ^a, hence the discrepancy in between 6,892 and 6,889
MRF ADMINISTRATION DATA		
901 Introduction date	13,253 1993 Sept163	This field is mandatory but it was not populated when MRF was created so 53,480 records are

CLASSIFICATION FIELDS		
MRF FIELD	**NUMBER OF RECORDS**	**COMMENTS**
	1993 Oct1 1994398 199546 1996541 1997919 1998134 19991,117 20002,290 20014,101 2002868 20031,307	missing the date. The field can be used to trace how many new numbers have been added in each annual update
903 Introduction source	12,785	This field should be mandatory but it turns out that the field was not populated when MRF was created and 53,948 are missing the source data.
904 Introduction Comments	698	
911 Cancellation date (yymm)	0	Every cancelled record is exported from the MRF and this field should be present in cancellation database only
912 Replacing UDC notation (if any)	0	This content is relevant for cancelled numbers only
913 Source of cancellation.	0	This content is relevant for cancelled numbers only
914 Cancellation comments	1	This field is relevant for cancelled numbers only. One record contains text by mistake
921 Last revision date	3,075	
922 Field(s) revised	3,067	
923 Source of revision	2,935	In 140 records with recorded revision date, source of revision is missing
924 Revision comments	477	
925 Revision history	273	
951 Index only UDC notations	566	Field is used to record UDC notation that is mentioned in notes and examples of combinations and is not searchable in any other way
952 Use special characters	330	If special character is used somewhere in the text , the word is repeated here with description of character that needs to be displayed
955 Editorial annotations	85	Field contains notes containing explanations necessary for record use, display or printing that is not scope or application note.
957 - Administrative notes	12	Contains notes relevant for revision and the sources used for revision
958 - For next E&C	0	This field is meant to be used by Editor in Chief to record future work. Field is empty as Editor in Chief does not use the MRF database
999 Temporary work	0	see comment above at 958

6.2 UDC MRF data element comparison with new MRF database proposal[166]

DATA ELEMENTS		
MRF 1993-2005	**MRF PROPOSAL 2003**	**COMMENTS**
001 UDC number (no subfields)	**001 UDC number** (no subfields)	No change introduced! (see also 004)
002 Table (no subfields) Codes a Table 1a – Coordination. Extension b Table 1b – Relation. Subgrouping. Order-fixing c Table 1c – Com. aux. of language d Table 1d – Com. aux. of form e Table 1e – Com. aux. of place f Table 1f – Com. aux. of ethnic grouping g Table 1g – Com. aux. of time h Table 1h –notations from non-UDC source i Table 1i – Com. aux. of point-of-view [!] k Table 1k – Com. aux. of general characteristics l Section II. Special auxiliary subdivisions M Main table	**002 Table** (no subfields) Codes a Table 1a – Coordination. Extension b Table 1b – Relation. Subgrouping. Order-fixing c Table 1c – Com. aux. of language d Table 1d – Com. aux. of form e Table 1e – Com. aux. of place f Table 1f – Com. aux. of ethnic grouping gTable 1g – Com. aux. of time h Table 1h –notations from non-UDC source i Table 1i – Com. aux. of point-of -view [!] k Table 1k – Com. aux. of general charact. l Section II. Special auxiliary subdivisions M Main table	No change introduced!
003 Special aux. type (no subfields) Codes: A hyphen (–) auxiliary B point-nought (.0) auxiliary C inverted comma (`) auxiliary D other, e.g. final digits (…1/…9)	**003 Type of special auxiliaries** (no subfields) Codes: A hyphen (–) auxiliary B point-nought (.0) auxiliary C inverted comma (`) auxiliary D other, e.g. final digits (…1/…9)	No change introduced!
004 Combination type (no subfields) Coded value list: a Plus addition [+] b Stroke range [/] c Simple number d Intercalation e Colon [:] combination f Double colon [::] combination g Combination with Com. Aux. of Lang. [=...] h Combination Com. Aux. of Form [(0...)] i Combination with Com. Aux. of Place [(1/9)] k Combination with Com. Aux. of Ethnic Grouping [(=...)] l Combination with Com. Aux. of Time ["..."] m Combination with non-UDC notation [*] n Combination with A/Z o Combination with .00 special auxiliary p Combination with -0 Com. Aux. of Persons/Materials q Combination with hyphen (-) spec. aux. r Combination with point-nought (.0) spec. aux. s Combination with apostrophe (') spec. aux.	**004 Combination type**. REPEATABLE (no subfields) Coded value list a Plus addition [+] b Stroke range [/] c Simple number d Intercalation e Colon [:] combination f Double colon [::] combination g Combination with Com. Aux. of Lang. [=...] h Combination Com. Aux. of Form [(0...)] i Combination with Com. Aux. of Place [(1/9)] k Combination with Com. Aux. of Ethnic Grouping [(=...)] l Combination with Com. Aux. of Time ["..."] m Combination with non-UDC notation [*] n Combination with A/Z o Combination with .00 special auxiliary p Combination with -0 Com. Aux. of Persons/Materials q Combination with hyphen (-) spec. aux. r Combination with point-nought (.0) spec. aux. s Combination with apostrophe (') spec. aux.	This element was suggested in 1993 but was not implemented. No change introduced!
	005 Hierarchical next higher notation. REPEATABLE Subfields ^w Notation ^l Language ^t Text note Instruction: If ^t is used ^l is mandatory	New element!
	006 Special auxiliaries, valid for the notation in field 001. REPEATABLE Subfields ^wNotation - begin of range ^x Notation - end of range ^l Language	New element!

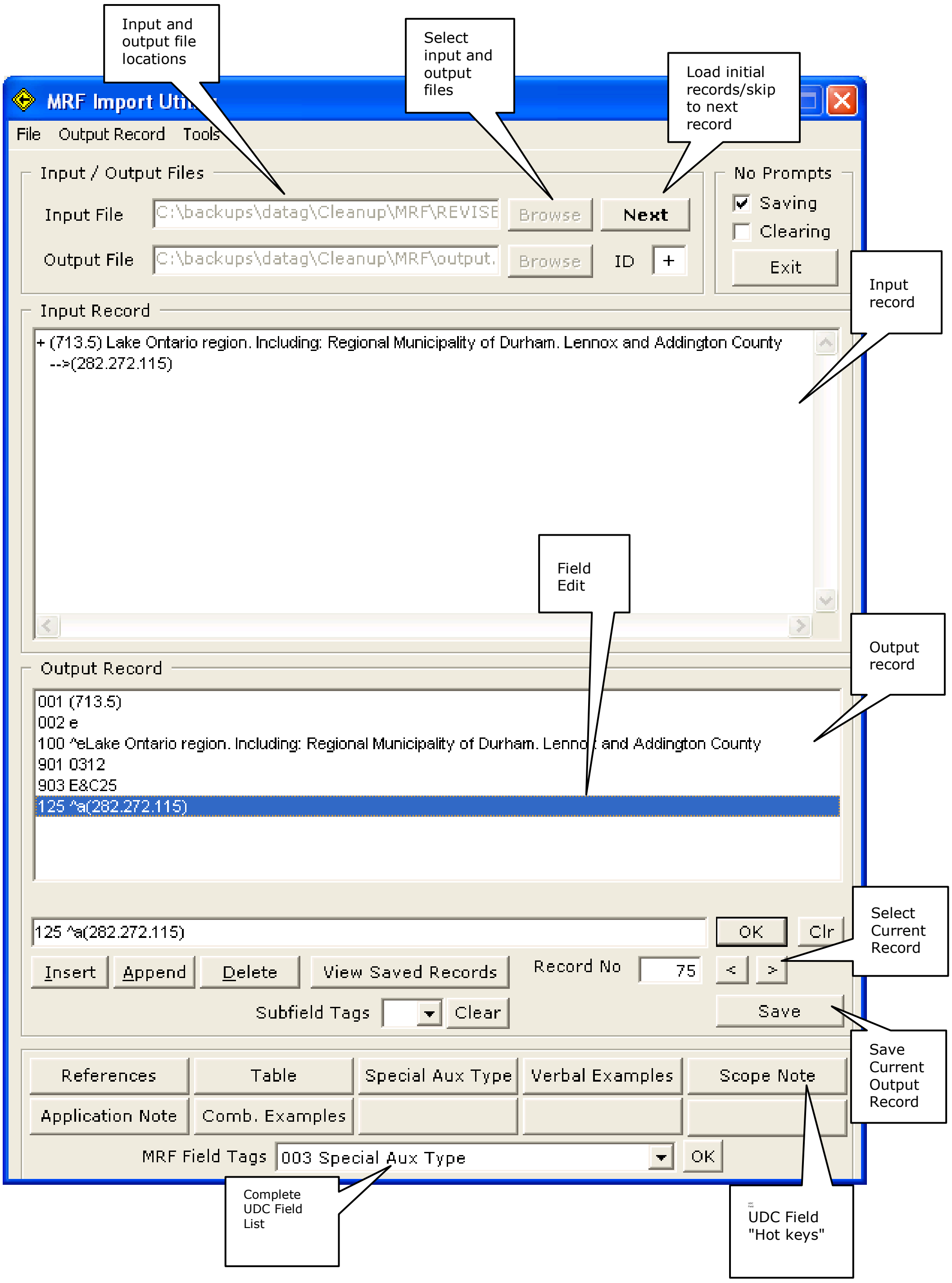
166 This proposal was put forward by G. Riesthuis and was published in E&C 25 (2003)

DATA ELEMENTS		
MRF 1993-2005	**MRF PROPOSAL 2003**	**COMMENTS**
	^t Text note	
010 Parallel derivation (no subfields)	**010 Parallel derivation** (no subfields)	No change.
011 Instruction for parallel division Subfields ^a Source notation for the parallel division ^b Target notation (if not identical with the number in field 001) ^t English annotation ^g German annotation	**011 Instruction for parallel division** Subfields ^wSource notation for parallel division (repeatable) ^x Target notation (if not identical with the notation in field 001) ^l Language ^t Text	Subfield change!
012 Special auxiliary supplied by parallel division (no subfields) Codes A hyphen (–) auxiliary B point-nought (.0) auxiliary C inverted comma (`) auxiliary D other, e.g. final digits (…1/…9)	012 **Special auxiliary supplied by parallel division** (no subfields) Codes A hyphen (–) auxiliary B point-nought (.0) auxiliary C inverted comma (`) auxiliary D other, e.g. final digits (…1/…9)	No change!
013 **Special auxiliary supplied by annotation** (no subfields) Codes Ahyphen (–) auxiliary B point-nought (.0) auxiliary C inverted comma (`) auxiliary D other, e.g. final digits (…1/…9)	013 **Special auxiliary supplied by annotation** (no subfields) Codes A hyphen (–) auxiliary B point-nought (.0) auxiliary C inverted comma (`) auxiliary D other, e.g. final digits (…1/…9)	No change!
100 Description Subfields ^eEnglish text ^gGerman text	**100 Description**. REPEATABLE Subfields ^l Language ^t Text ^n Note	Subfield change
105 Verbal examples Subfields ^eEnglish text ^gGerman text	**105 Verbal description**. REPEATABLE Subfields ^l Language ^t Text ^n Note	Subfield change, element name change!
110 Scope note Subfields ^eEnglish text ^gGerman text	**110 Scope note**. REPEATABLE Subfields ^l Language ^t Text ^n Note	Subfields change!
111 Application note Subfields ^eEnglish text ^gGerman text	**111 Application not**e. REPEATABLE Subfields ^l Language ^t Text ^n Note	Subfield change!
115 Combination examples Subfields: ^aDirect addition ^bColon (:) combination ^cFull notation Used if the first component differs from the notation in field 001. ^rReference(s) to UDC-number ^dEnglish description ^gGerman description ^nEnglish annotation ^yGerman annotation	**115 Combination examples**. REPEATABLE Subfields: ^xColon combination ^y Direct addition ^zFull notation (if the first component differs from the notation in the field 001) ^wReferences ^l Language ^t Text ^n Note	Subfield change!
120 Parallel division examples Subfields ^a Notation ^dEnglish description ^gGerman description	**120 Parallel division examples**. REPEATABLE Subfields ^w Notation ^zReferences ^l Language ^t Text ^n Note	Subfield change!

DATA ELEMENTS		
MRF 1993-2005	**MRF PROPOSAL 2003**	**COMMENTS**
125 References Subfields ^a Notation ^tEnglish annotation ^gGerman annotation	**125 References**. REPEATABLE Subfields ^w Notation ^l Language ^t Text ^n Note	Subfield change!
	126 Index terms. REPEATABLE Subfields ^w Notation (if different from notation in field 001, e.g. from a 120 field) ^l language ^t text	New element! It is not clear why notation is planned for this element, which should contain simple uncontrolled keywords for whatever notation is in field 001.
MRF ADMINISTRATION DATA		
901 Introduction date	**901 Introduction date** Subfields ^xdate (former field 901) ^zsource (former field 903) ^l Language ^tComment (former field 904)	Subfield change!
903 Introduction source	**Replaced by 901**	Field removed!
904 Introduction Comments	**Replaced by 901**	Field removed!
911 Cancellation date (yymm)	**911 Cancellation date** Subfields ^xdate (former field 911) ^yreplacing UDC notation (if any) (former field 912) ^zsource (former field 913) ^l Language ^nComment (former field 914)	Subfield change!
912 Replacing UDC notation (if any)	Replaced by 911	Field removed!
913 Source of cancellation	Replaced by 911	Field removed!
914 Cancellation comments	Replaced by 911	Field removed!
921 Last revision date	Replaced by 925	Field removed!
922 Field(s) revised. Identified by field number; if more fields are affected, numbers should be separated by % (percent sign).	Replaced by 925	Field removed!
923 Source of revision	Replaced by 925	Field removed!
924 Revision comments	Replaced by 925	Field removed!
925 Revision history	**925 Revision history** (repeatable) Subfields ^xDate (replacing field 921) ^yField (replacing field 922) ^zSource(replacing field 923) ^l Language ^tText (replacing field 924)	Subfields change!
951 Index only UDC notations	**951 Term for index** (repeatable) Subfields ^w notation ^l Language ^t Text	Subfield change!
952 Use special characters	**952 Special characters** (repeatable) Subfields ^l Language ^t Text	Subfield change!

DATA ELEMENTS		
MRF 1993-2005	**MRF PROPOSAL 2003**	**COMMENTS**
955 Editorial annotations	**955 Edition annotation** Subfields ^l Language ^t Text	Subfield change!
957 - Administrative notes	**957 Administration note** Subfields ^l Language ^t Text	Subfield change!
958 - For next E&C	**958 Next E&C** Subfields ^l Language ^t Text	Subfield change!
999 Temporary work	**999 Temporary work** Subfields ^l Language ^t Text	Subfield change!

6.3 UDC MRF Scanner Utility Program



The MRF Scanner is used to read a plain text file from the annual E&C publication, encoding the fields and producing a file that can be processed by the FANGORN2 program, which in turn produces an ISO2709 formatted file for import into the UDC MRF database.

MRF Scanner speeds up the process by automatically structuring text and sets default data such as introduction date and source, table, inserting tags between records and in front of main data elements. The parts of the text that cannot be encoded automatically are application/scope notes and subfields in examples of combinations and references. In this case MRF Scanner allows for controlled checking of UDC numbers (one by one) and insertion of necessary codes using 'hot keys'.

The program runs in a Microsoft Windows environment and was written using Microsoft Visual Basic (Overfield, 2001).

The first version of the program was written and used tested in the MRF update in 2000 and the program was subsequently updated and handed to the UDCC office in The Hague, together with a user manual.

If the source data file were presented in a more structured manner, the program would be able to automate fully the process for upload to the UDC MRF database.

6.4 UDC MRF licence categorisation

The UDC Consortium's categorization of the UDC MRF licences, illustrated here, was put together in 2003 and made publicly available for the first time in the *Extensions &Corrections to the UDC, 26* distributed in January 2005.

The Consortium differentiates prices for the following six UDC MRF licences[167]:

- **Publishing**: 1300 euro p.a. (i.e. for company or organization business use
- **User**: 650 euro p.a. (i.e. for individual business use)
- **Experimental**: 650 euro (normally issued for up to 1 year only)
- **Internet**: 3150 euro p.a. (i.e. for use on a public network)
- **Intranet**: 1575 euro p.a. (i.e. for use on a private, secure network)
- **Consumer**: 130 euro p.a. (i.e. for private interest or use) publishing

(How to obtain the UDC?, 2004: 46)

From this categorisation it appears that the Consortium has confused the following:

(i) the use of the UDC **system** with the use of the UDC MRF

(ii) the type of use of UDC MRF with the type of technology/medium

For instance, the Consortium is interested in whether or not the UDC MRF is **used** on the *Internet* or *intranet* but does not provide different pricing if the UDC MRF is **published** on the Internet or Intranet. It is not clear why would Internet/intranet technology apply to **use** and not to publishing, experimental use, or consumer use.

It is not clear what is the difference between a *user licence* and an *intranet user licence*. The Consortium, obviously, anticipates that there is a kind of individual business user who does not use the intranet, extranet or any other kind of local area network. If this is the case, and if 'individual business users' actually means one

---

[167] Licences are normally issued and paid for a period of three years during which a user will receive three different UDC MRF files. In these three annual files, database record numbers will be changed and UDC numbers may be cancelled and re-used with different meaning. Essentially this means that only one of these three different files will be used in a real-life application and from the point of online use the three-year licence may not make any sense. This is especially true because of the fact that UDC data, when distributed, does not contain a unique number identifier that would enable some kind of automation or control of UDC updates in target systems, even when users invest in a middleware solution.

personal computer, then the price should be linked to the number of computers or workstations/terminals on which the UDC MRF is used. It is not also clear why the Consortium confuses security/limitation of use with the type of network, i.e. Internet/intranet. Any network could potentially be secure and closed or insecure and open.

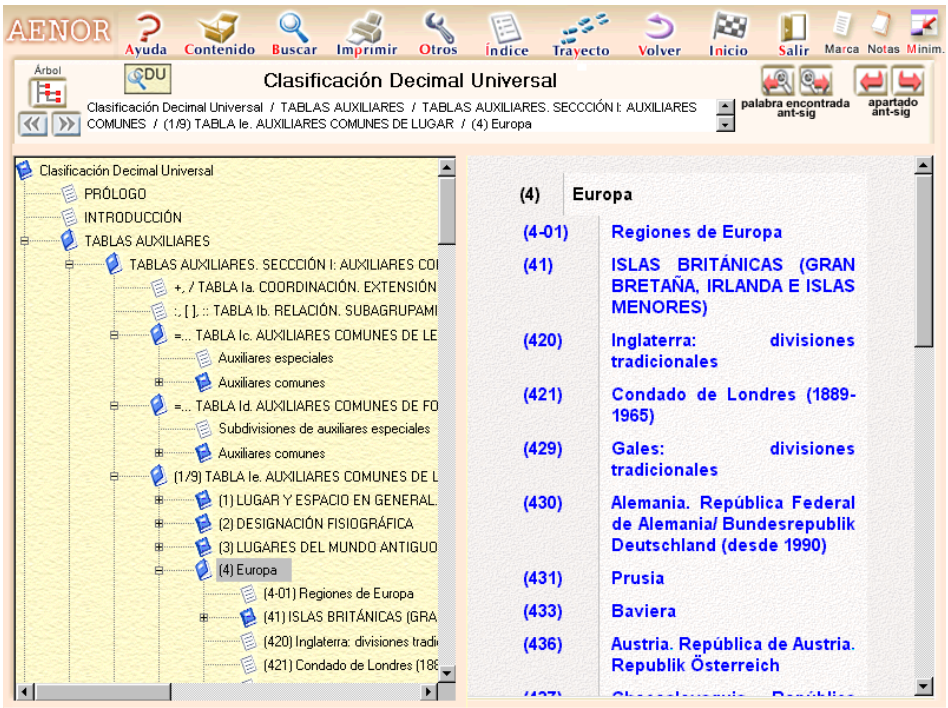
But most importantly, what is considered to be 'UDC use'? There is no explanation as to what the **user licence** covers: the use of the 'UDC MRF' or the use of the 'UDC system'. Logically, one would expect that this must be related to the UDC MRF because this is to what the price listed above refers. The problem is that one cannot **use** the UDC MRF on the Internet. One can **use** the UDC **system** on the Internet/intranets but it is not possible to use the UDC MRF database **file** on the *Internet/intranets* as this would immediately become something else: **publishing** and, therefore, would be regulated through a **publishing licence** and not through a **user licence.** Unless, and perhaps this is the only explanation, the UDC Consortium in this case speaks about the cost of **UDC system** use and not **UDC MRF** use.

Another issue is that a copyright declaration or agreement is not distributed as part of the UDC MRF file and there is no publicly available copyright declaration by the UDC Consortium that might clarify the difference between the use of the **UDC system** and the use of the **UDC MRF**. This is extremely important with respect to existing users worldwide. The fact is that the UDC ownership changed in 1992 and users may not be entirely sure whether they have to pay for the use of the UDC system or whether they have to pay for the *product* when they use it and are free to use the *system* for any application and collection size. If the use of the UDC system (and not the UDC MRF) is copyright protected, then the UDC Consortium has to clarify what this means and how they price this for thousands of institutions worldwide who use the system but not the UDC MRF.

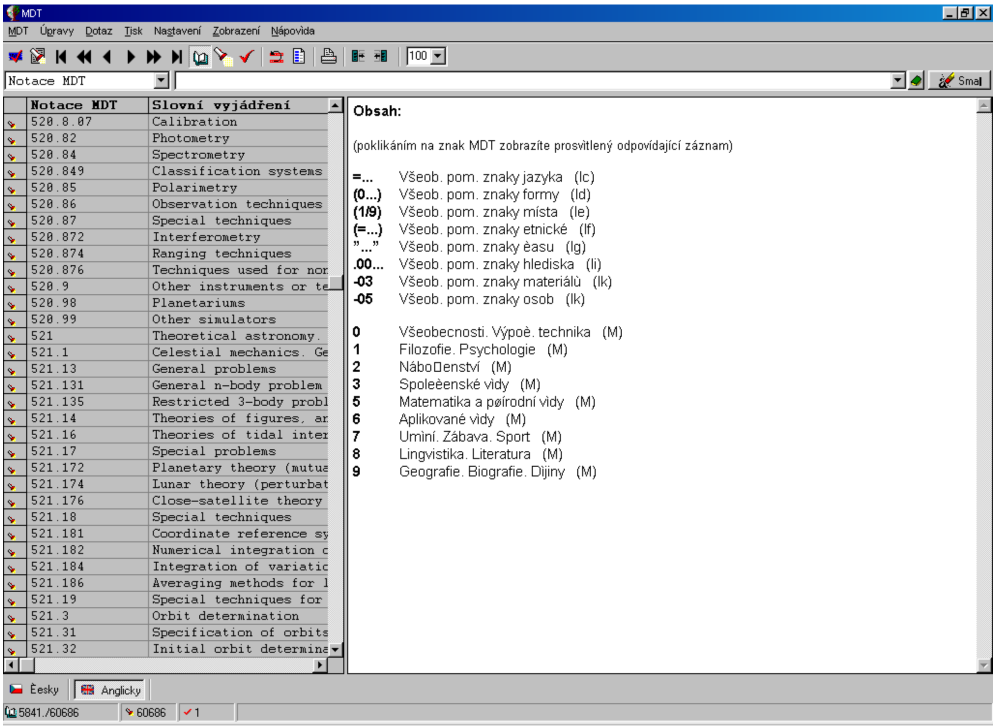
7. UDC CLASSIFICATION TOOLS

7.1 CD-ROM tools

7.1.1 Spanish UDC on CD-ROM



7.1.2 Czech CD-ROM



7.2 Schedules on the Web

7.2.1 UDC-online 2001-2004

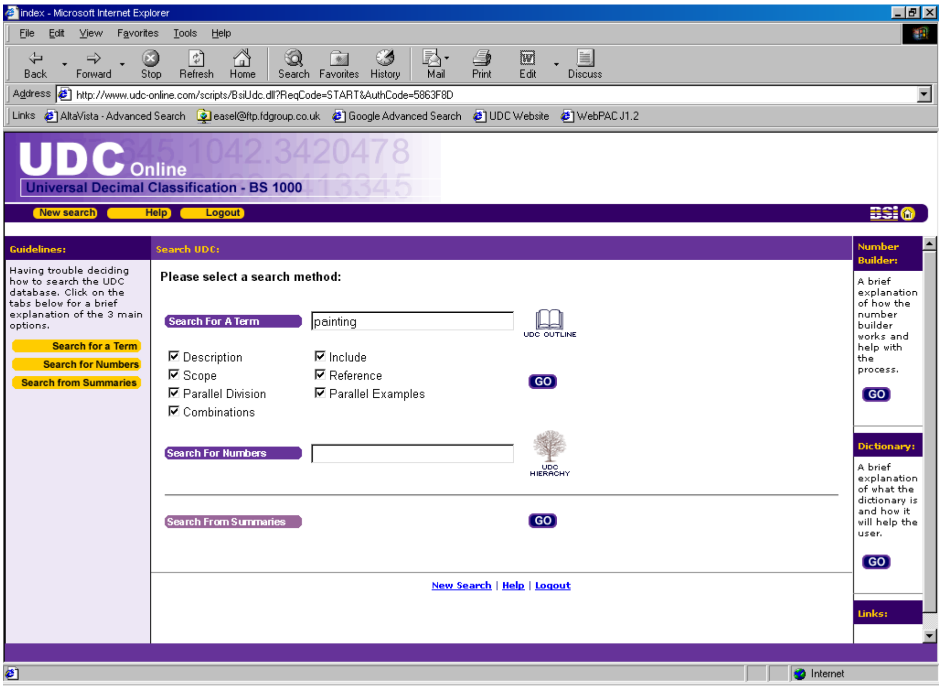


Figure 1.3 Main search page (http://www.udc-online.com)

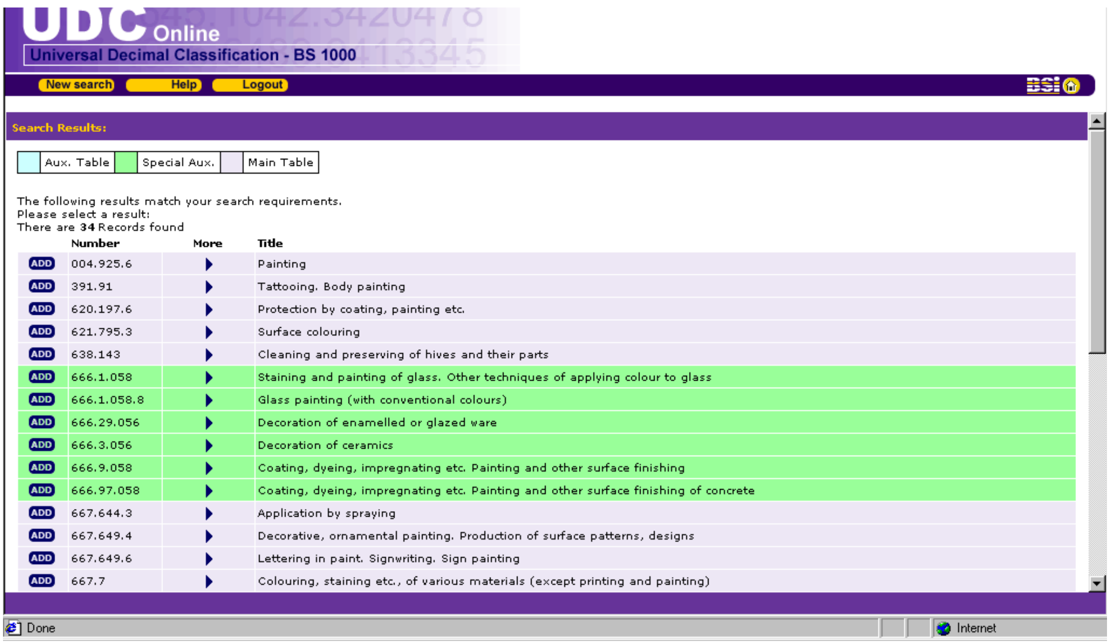


Figure 1.4 Search result display

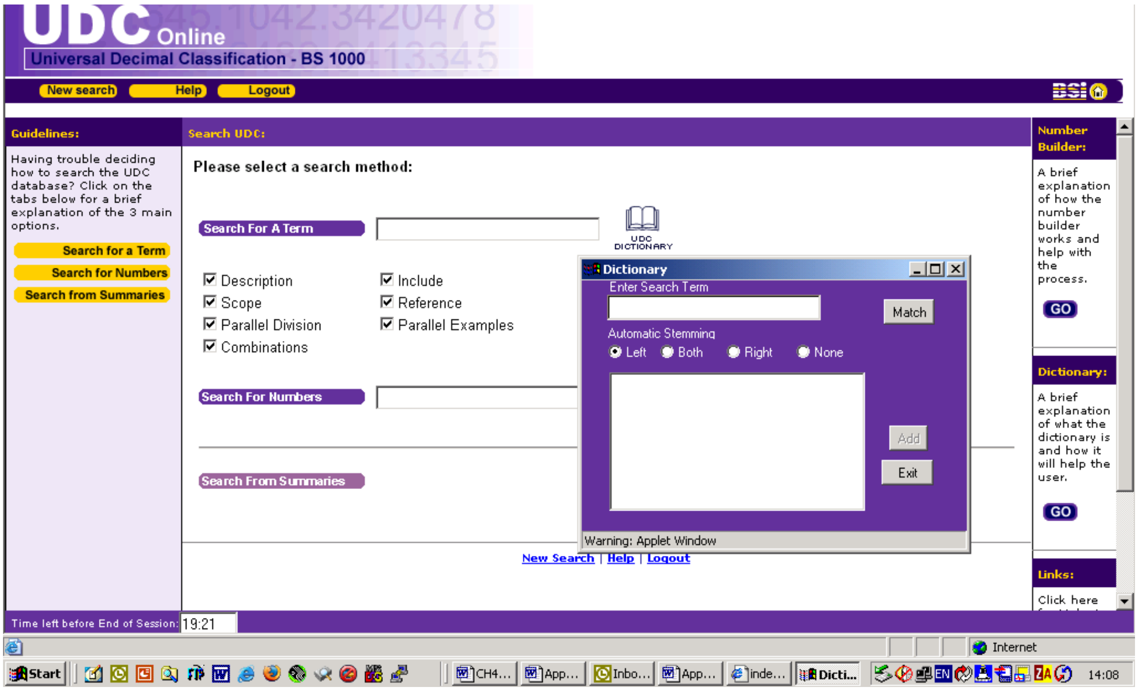


Figure 1.5 Option of search term selection from the 'dictionary'

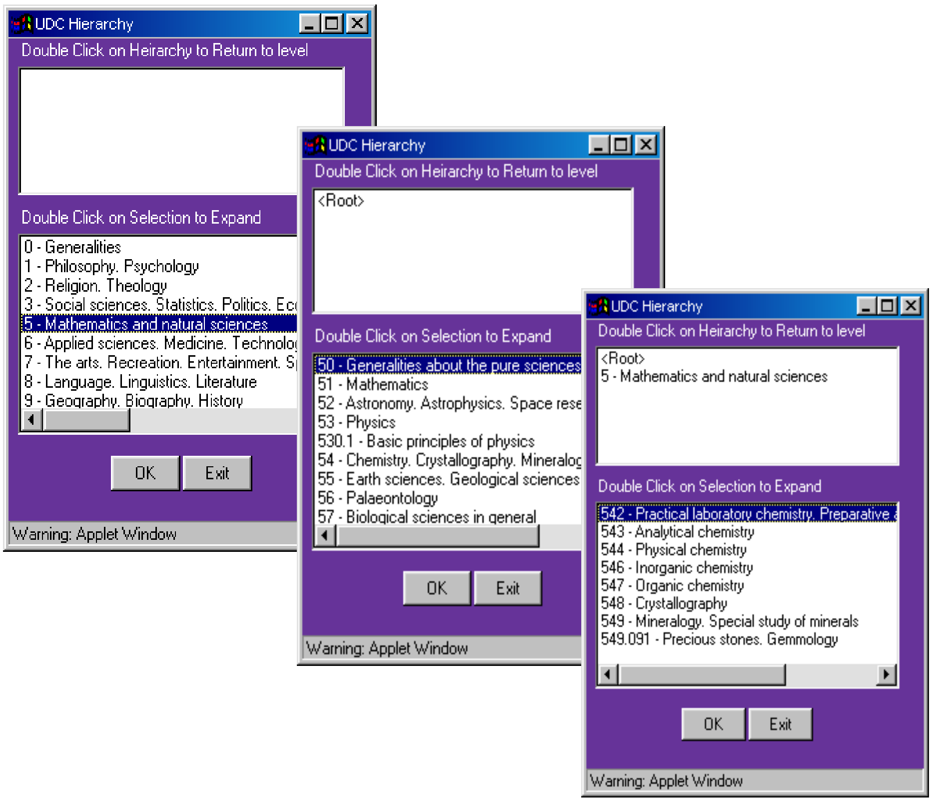


Figure 1.6 UDC-online top/down hierarchy browsing

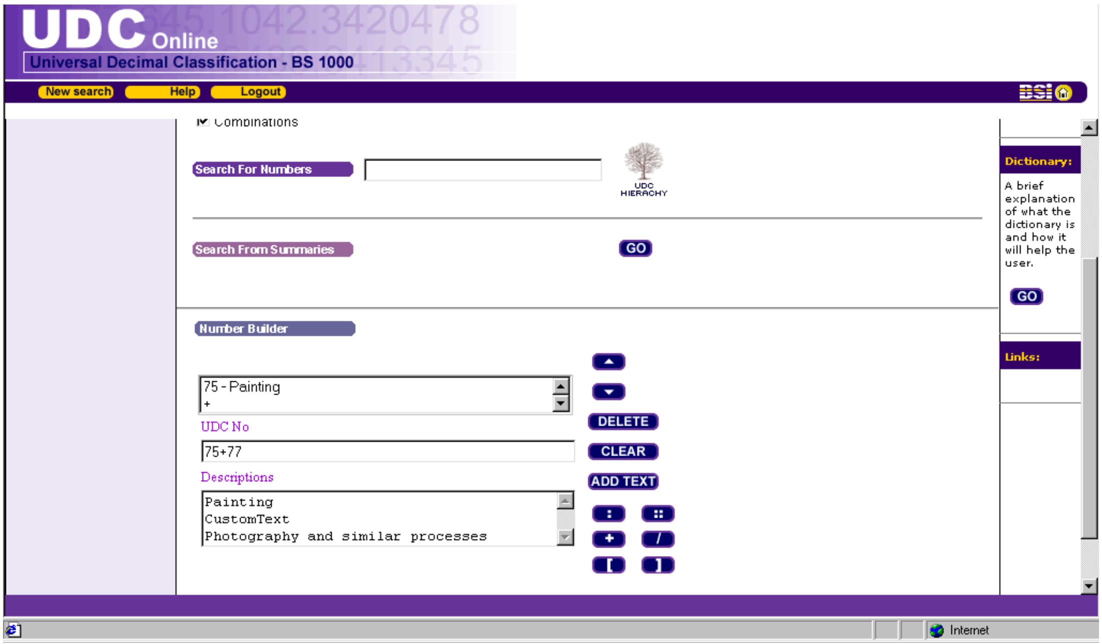


Figure 1.7 Number builder function

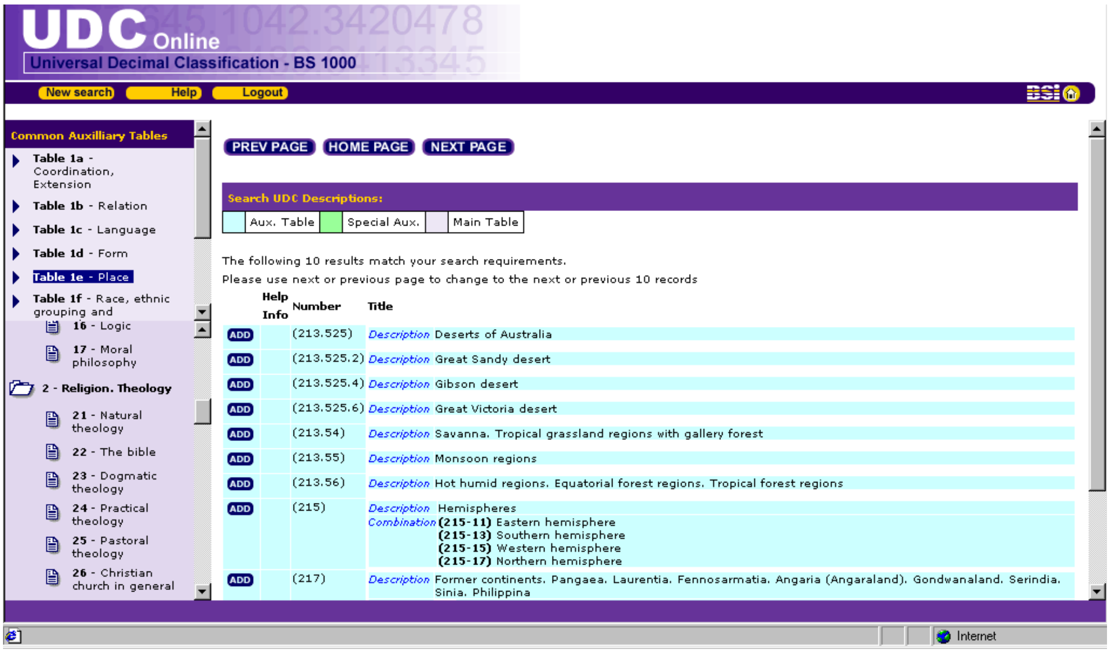


Figure 1.8 Browsing UDC tables

7.2.2 Czech UDC on the Web

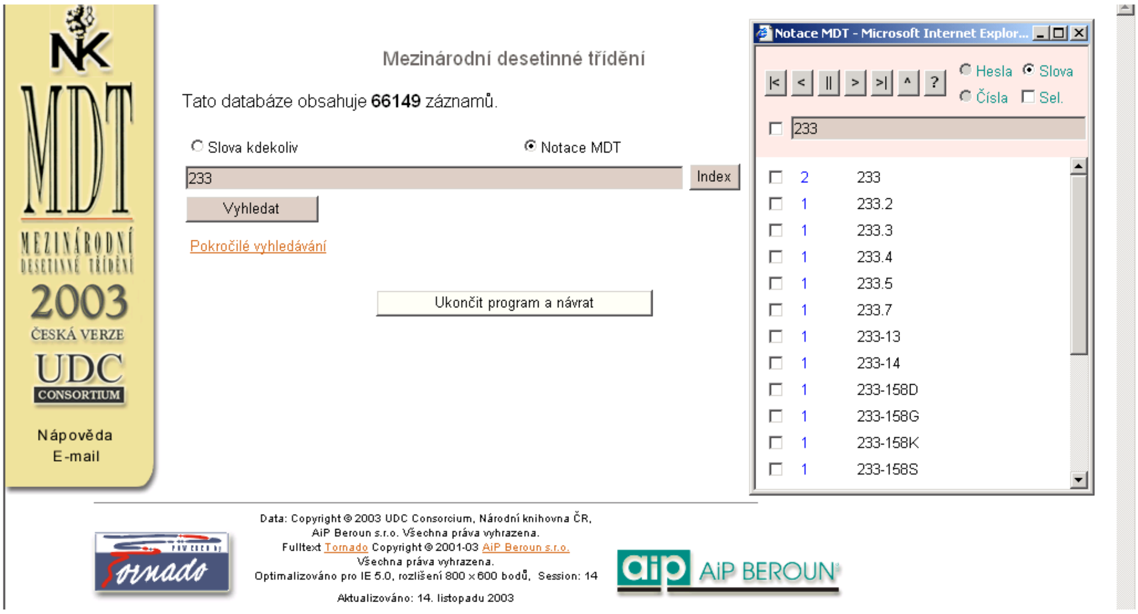


Figure 1.9 Main search page (http://aip.nkp.cz/mdt/)

Figure 1.10 Search result

7.2.3 Swedish UDC on the Web[168]

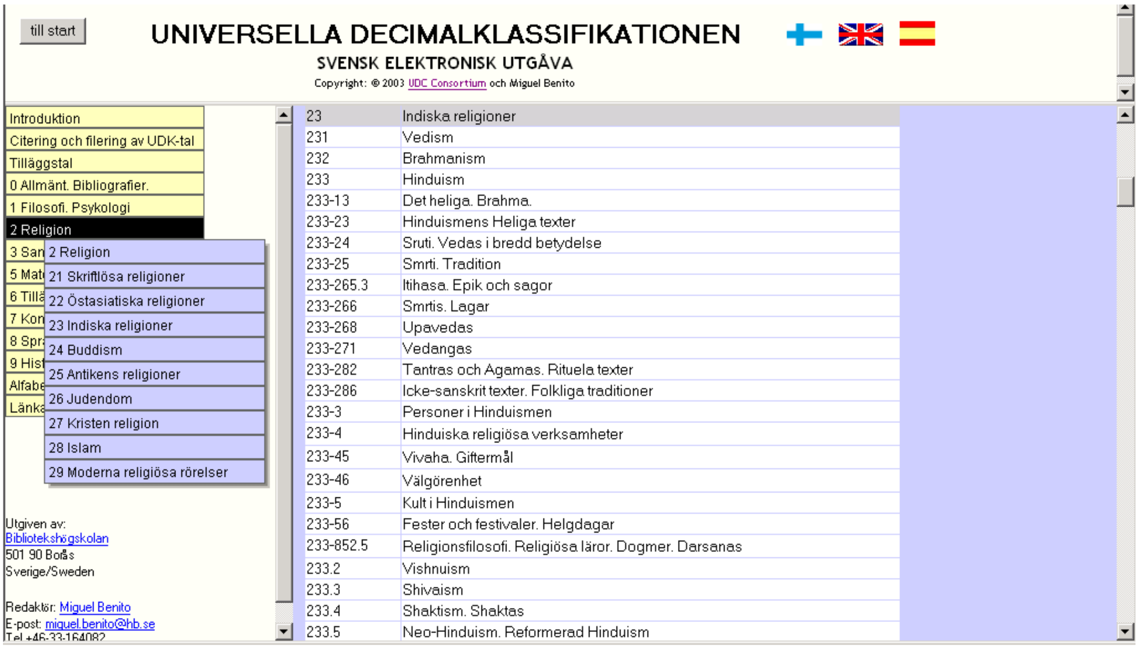


Figure 1.11 Browsing Swedish UDC (http://www.hb.se/bhs/udk/)

7.2.4 Italian UDC on the Web[169]

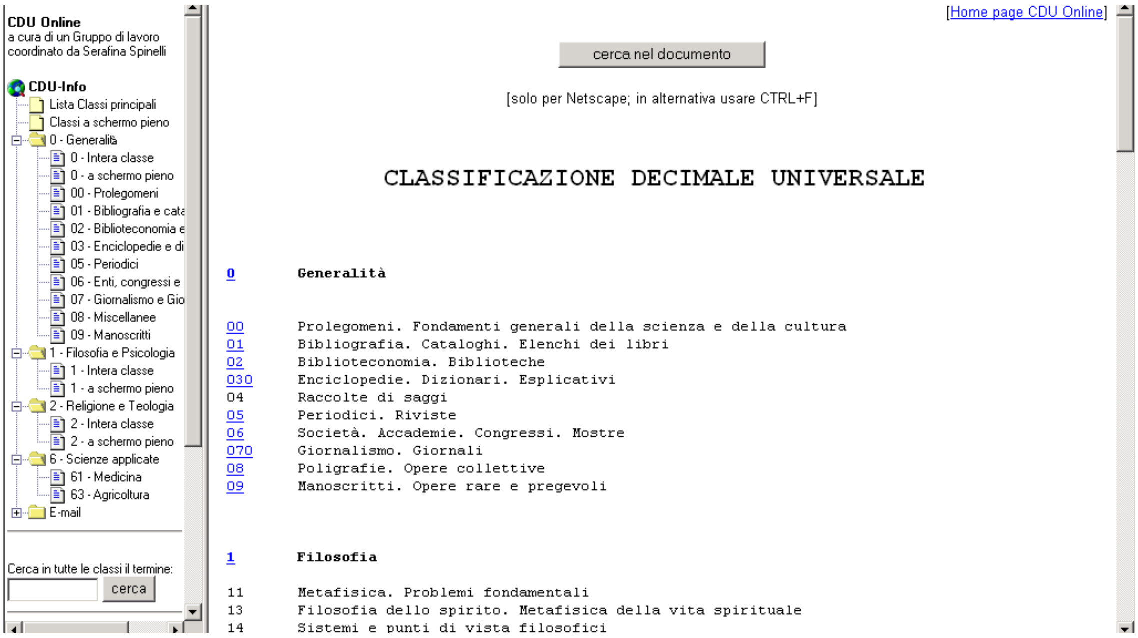


Figure 1.12 Abridged Italian UDC (http://mail.biocfarm.unibo.it/%7Espinelli/cdu/)

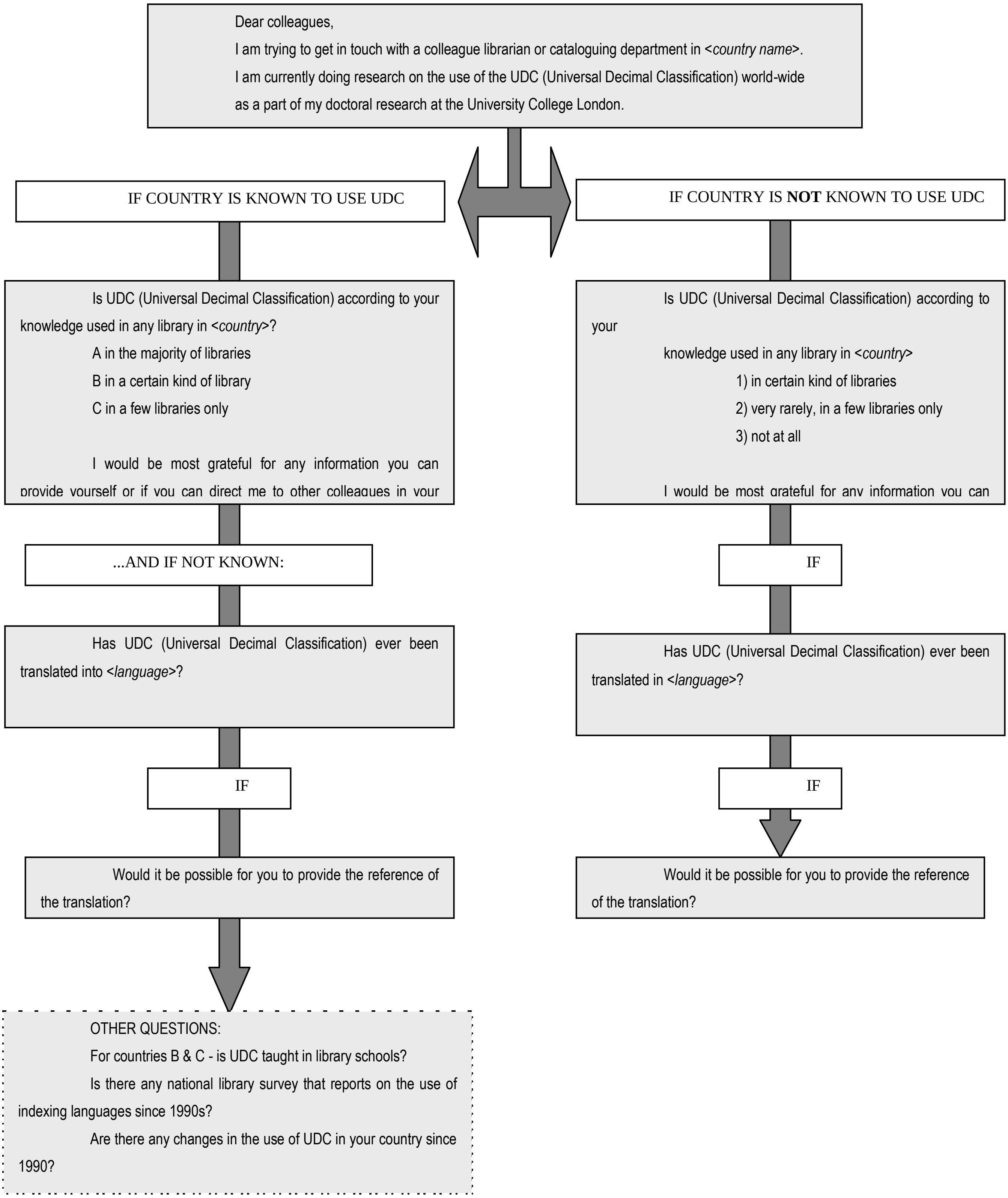
---

[168] This classification tool contains a selection of numbers from UDC MRF. It contains notations and descriptions and provides only browsing function. Outline in Finnish and Spanish are available.

[169] This offers browsing/searching of an abbreviated Italian edition from 1987.

8 UDC USERS SURVEY

8.1 E-MAIL INTERVIEW DESIGN DIAGRAM



8.2 Samples of initial e-mail interview messages

FRENCH

Madame, Monsieur,

Je suis à la recherche de renseignements concernant la Classification Decimale Universelle (CDU) à l'University College London.

J'essais de contacter des specialistes de bibliothèque afin de trouver des nformations sur l'utilisation de ce systeme dans votre pays.

Je n'ai que une question tres simple: Est-ce que la CDU est-elle utillisee en <....> ?
Pourriez vous m'envoyer toute information susceptible de m'aider dans ma recherche (d'autres contacts par exemple)?
En vous remerciant à l'avance, je vous prie de croire, Madame, Monsieur, à l'expression de mes meilleures salutations...

SPANISH

Queridos colegas,

Estoy intentando contactar con bibliotecarios o departamentos de catalogacion en Cuba. Estoy haciendo my proyecto de tesis sobre la CDU (Clasificacion Decimal Universal), aqui en el University College London, y necesito saber si en su pais usan la CDU.
Estaria extremadamente agradecida si pudieran contactarme o darme el email de alguien que pueda servirme de ayuda - en caso de que usted no este familiarizado con el tema

Muchas gracias por avanzado.
Sinceramente...

PORTUGUESE

Caros Colegas

Estou a tentar contactar bibliotecarios ou departamentos de catalogacao em <...>, para obter informacao sobre o uso da CDU (Clasificacao Decimal Universal), para um projecto de investigacao aqui no University College London.
A CDU é usada actualmente nas bibliotecas do seu país?

Muito agradecia se pudesse colocar-me em contacto com alguem que me possa ajudar.

Agradecendo desde ja antecipadamente a ajuda que possa dar.
Melhores cumprimentos...

8.3 Samples of information obtained through e-mail interview

8.3.1 Examples of a few informative and high quality responses

ZAMBIA
e-mail message received 06/08/2004

Dear Aida,

I have contacted all people involved in documentation centres and there is no body who is using UDC. Apparently, we are having a meeting for all librarians, so, I could not miss any one.

Chrispin Hamooya
National Archives of Zambia [naz@zamnet.zm]

MOZAMBIQUE
e-mail message received 08/09/2004
Cara colega,

As minhas desculpas pelo atraso da resposta à sua mensagem de 2 de Agosto mas o sistema informático do CIDOC esteve parado por vários motivos e só agora foi reposto; portanto só agora tivémos acesso à mensagem. Esperando que a informação ainda seja útil:

Entre Janeiro de 2001 e Maio de 2002, ao abrigo de um projecto do Fundo Bibliográfico de Língua Poretuguesa, de levantamento e caracterização das unidades documentais existentes no país (arquivos, bibliotecas e centros de documentação e informação), foi efectuado um inquérito em todo o país, abrangendo 261 unidades documentias, que se considera representarem cerca de 90% das existentes em MZ. 114 das 261 unidades inquiridas utilizam a Classificação Decimal Universal (CDU) embora, na maioria dos casos, usem pouco mais do que as 10 classes principais.
Com base nos inquéritos às unidades documentais foi criada uma base de dados informatizada incluindo a descrição dos seus fundos e dos serviços que prestam e do pessoal ligado ao processamento técnico e atendimento público.

Como produto dessa base de dados, em Outubro de 2003 foi publicado pelo Fundo o "Directório dos arquivos, bibliotecas e centros de documentação e informação existentes em Moçambique".
Teremos o maior gosto em oferecer-lhe um exemplar desde que, pelo seu lado, consiga um portador para o mesmo. O CIDOC - Instituto Médio de Ciências Documentais (instituição subsidiária do Fundo Bibliográfico de Língua Portuguesa), única escola existente no país para a formação de pessoal (básico e médio) para as áreas de arquivo, biblioteca e centro de documentação, criado em 1998, ministra, na disciplina de Classificação, a teoria e prática da CDU, como estudo de caso.

Esperando que a informação responda às suas perguntas, aceite as minhas mais cordiais saudações.

Wanda do Amaral

Assessora do Fundo Bibliográfico de Língua Portuguesa, Presidente da Comissão Instaladora do CIDOC

SENEGAL
e-mail message received 23/07/2004

Salut,
La CDU est bien utilisee et meme enseignee. Au niveau de notre centre de documentation nous utilisons le systeme de classification SATIS
(Socially Appropriate Technology International Information Services). Je vous donne l'adresse de l'Ecole des Archivistes Bibliothecaires et Documentalistes de lUniversite de Dakar qui pourront vous orienter (ebad@ebad.ucad.sn).
Cordialement,
Emmanuel Seck
Enda Energie
BP 3370 Dakar
energy2@enda.sn

e-mail message received 23/07/2004

La CDU est bien utilisée au Sénégal. Mais je ne suis pas au courant de l'existence d'une traduction en Arabe. Elle est utilisée tant au niveau de la Bibliographie du Sénégal qu'au niveau de certaines bibliothèques.
La bibliographie du Sénégal est vendue à 30$ US : 25$ pour un numéro et 5$pour les frais d'envoi pour un numéro.
Nous sommes à votre disposition pour l'acquisition de la bibliographie du Sénégal en cas de besoin.

Bonne étude.
Diallo [bdas@primature.sn]

ANGOLA
e-mail message received 20/07/2004

Prezada Sra. Slavic,
A CDU é utilizada actualmente nas bibliotecas angolanas, fundamentalmentenas públicas, na nacional e nas universitárias.
Estou ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,
Maria José Ramos
Biblioteca Nacional [bibliotecanacional@netangola.com]

MALI
e-mail message received 03/09/2004

Bonjour Aida

Je suis heureux de recevoir votre émail. Pour répondre à votre quetion je dirais que la CDU est utilisée au Mali à la bibliothèque Nationale et dans les bibliothèques universitaires (ENSUP, ENA,ENI, Ecole de Medecine). Pour des contacts : Mme Koumaré Fatoumata Koumaré,émail : koumarefatoumata2003@yahoo.fr et M. Amadou Bekaye SIDIBE : absidibe@hotamail.com peuvent vous fournir des informations.

Si vous avez besoin d'autres informations n'hésiter pas à me demander. A bientôt.

Lamine CAMARA

Direction Nationale des Archives du Mali

Secrétaire général de l'Association Malienne des Bibliothécaires Archivistes et Documentalistes

Complexe Bibliothèque Natioanale/Archives Nationales Hamdallaye ACI 2000 BP.159 Bamako (Mali)

8.3.2 A few examples of least informative answers

HONG KONG
e-mail message received 20/08/2004

Dear Ms Slavic,
I have no idea about which library in Hong Kong uses UDC. Sorry.
Regards
Wai King
wkling [wkling@hkapa.edu]

MALI
e-mail message received 04/08/2004

bonjour Madame,
C'est quoi la CDU s'il vous plait?

Aguibou COULIBALY
Chargé d'Informatique de Maintenance et Réseau
Institut du Sahel - INSAH / CILSS
BP:1530, Bamako, Mali

MEXICO
e-mail message received 03/09/2004

Estimada Aída,

Le comento que en México son pocas las instituciones que utilizan el CDU, por el momento no le puedo precisar cuantas y a quien contactar, pero en cuanto cuente con los datos se los haré llegar.

Saludos

Ariel Rodriguez [ariel@cuib.laborales.unam.mx]

BOLIVIA
e-mail message received 26/07/2004

Estimada amiga
Nosotros trabajamos con CDD.
Atentamtne Ing. Freddy Chavez
CDIMA [cdima-cni@bolivia-industry.com]

8.4 UDC users survey data

Table 1.7 Preliminary data on countries using UDC

	IUFRO GFDC SURVEY[170]	UDC CONSORTIUM SOURCES			UDC MAILING LIST
		AENOR[171]	UDCC USER LICENCE	USERS ON THE UDCC WEB	
EUROPE	Finland, Italy, Norway, Switzerland, UK	Andorra, Italia Portugal, Spain, Sweden	Armenia, Belgium, Croatia, Czech Rep., Denmark, Finland, France, Germany, Hungary, Ireland, Italy, Latvia, Lithuania, The Netherlands, Poland, Portugal, Romania, Russia, Slovakia, Slovenia, Sweden, Switzerland, UK, Ukraine, Yugoslavia	Armenia, Belgium, Croatia, Czech Rep., Finland, Germany, Hungary, Ireland, Latvia, Portugal, Romania, Slovakia, Slovenia, Switzerland, The Netherlands, UK, Ukraine,	Belgium, Bosnia& Herzegovina, Croatia, Estonia, Finland, France, Georgia, Iceland, Ireland, Latvia, Norway, Portugal, Slovakia, Slovenia, Spain, Sweden, The Netherlands, UK
ASIA	India, Indonesia, Israel, Malaysia, Sri Lanka		China (Macao), India, Japan	India, Israel, Japan, Sri Lanka	India, Israel, Syria,
AFRICA		Tunisia		Senegal, Tanzania	South Africa
NORTH AND CENTRAL AMERICA	Canada	USA (Florida)	USA	Cuba, Guatemala, Trinidad & Tobago, USA	Canada, USA
SOUTH AMERICA		Argentina, Brazil, Chile, Cuba Guatemala, Mexico, Peru Uruguay	Brazil	Argentina, Peru, Uruguay,	Brazil,
AUSTRALIA AND SOUTH PACIFIC	Australia, Fiji			New Zealand, Australia	Australia
	13	10	30	32	26
		51 different countries (by UDCC)			
TOTAL	63 different countries				

---

170 In 2003, IUFRO unit on Global Forest Decimal Classification conducted a survey on the use of classification in special forestry libraries world-wide. Data collected included libraries using UDC.

171 AENOR is a Spanish UDC Consortium member who provided list of countries they sell their Spanish editions to. Data from 2001 and 2004 is used.

Table 1.8 List of countries included in the 2004 survey[172]

EUROPE	ASIA	AFRICA	NORTH AND CENTRAL AMERICA	SOUTH AMERICA	AUSTRALIA AND SOUTH PACIFIC
1. Albania	Afghanistan	Algeria	Antig & Barb	Argentina	Australia
2. Andorra	Armenia	Angola	Aruba	Bolivia	Fiji
3. Austria	Azerbaijan	Benin	Bahamas	Brazil	Fr. Polynesia
4. Belarus	Bahrain	Botswana	Barbados	Chile	Kiribati
5. Belgium	Bangladesh	Burkina F.	Belize	Colombia	Marshall Islands
6. Bosnia & H	Bhutan	Burundi	Bermuda	Ecuador	Micronesia
7. Bulgaria	Brunei	Cameroon	Canada	Fr. Guyana	Nauru
8. Croatia	Burma (Myanmar)	Cape Verde	Cayman Isl.	Guyana	New Caledonia
9. Cyprus	Cambodia	Centr.Afr.R.	Costa Rica	Paraguay	New Zealand
10. Czech R	China	Chad	Cuba	Peru	Niue
11. Denmark	Comoros	Congo	Dominica	Surinam	Palau
13. Estonia	Georgia	Congo DR	Dominican R.	Uruguay	Pap. New Guin.
14. Faroe Isl.	Hong Kong	Djibouti	El Salvador	Venezuela	Samoa
15. Finland	India	Egypt	Grenada		Solomon Isl.
16. France	Indonesia	Eq. Guinea	Guadaloupe		Tonga
17. Germany	Iran	Eritrea	Guatemala		Tuvalu
18. Greece	Iraq	Ethiopia	Haiti		Vanuatu
19. Hungary	Israel	Gabon	Honduras		
20. Iceland	Japan	Gambia	Jamaica		
21. Ireland	Jordan	Ghana	Martinique		
22. Italy	Kazakhstan	Guinea	Mexico		
23. Latvia	Kuwait	Guinea Bis	Neth. Antiilles		
24. Liechtenstein	Kyrgyzstan	Ivory Coast	Nicaragua		
25. Lithuania	Laos	Kenya	Panama		
26. Luxembourg	Lebanon	Lesotho	Puerto Rico		
27. Macedonia	Macao	Liberia	St Kitts & N.		
28. Malta	Malaysia	Libya	St Lucia		
29. Moldavia	Maldives	Madagascar	St Vinc. &Gre		
30. Monaco	Mongolia	Malawi	Trinid &Tob		
31. Netherlands	Nepal	Mali	USA		
32. Norway	North Korea	Mauritania			
33. Poland	Oman	Mauritius			
34. Portugal	Pakistan	Morocco			
35. Romania	Palestine	Mozambique			
36. Russian Fed.	Philippines	Namibia			
37. San Marino	Qatar	Niger			
38. Serbia &MN	Saudi Arabia	Nigeria			
39. Slovakia	Singapore	Reunion			
40. Slovenia	South Korea	Rwanda			
41. Spain	Sri Lanka	Sao Tome &			
42. Sweden	Syria	Senegal			
43. Switzerland	Taiwan	Seychelles			
44. U.K.	Tajikistan	Sierra Leone			

[172] The term "country" is not taken here in a strict geo-political sense, i.e. dominions, overseas departments etc. were taken as separate entities.

EUROPE	ASIA	AFRICA	NORTH AND CENTRAL AMERICA	SOUTH AMERICA	AUSTRALIA AND SOUTH PACIFIC
45. Ukraine	Thailand	Somalia			
46. Vatican City	Turkey	South Africa			
	Turkmenistan	Sudan			
	UAE	Swaziland			
	Uzbekistan	Tanzania			
	Vietnam	Togo			
	Yemen	Tunisia			
		Uganda			
		Zambia			
		Zimbabwe			
46	50	53	30	13	17

Table 1.9 Countries with unconfirmed status in using UDC

SURVEYED COUNTRIES (208)	**NO INFORMATION GATHERED IN 2004**	
EUROPE	5	Cyprus, Liechtenstein, Monaco, San Marino, Vatican City
ASIA	15	Afghanistan, Bahrain, Brunei, Cambodia, Comoros, Iraq, Jordan, Kuwait, Laos, Maldives, Mongolia, North Korea, Qatar, Saudi Arabia, Yemen
AFRICA	19	Burundi, Central African Republic, Chad, Congo, Cote d'Ivoire, Djibouti, Equatorial Guinea, Eritrea, Gabon, Guinea, Lesotho, Liberia, Mauritania, Reunion, Rwanda, Sao Tome & Principe, Seychelles, Somalia, Swaziland
NORTH AND CENTRAL AMERICA	6	Bermuda, El Salvador, Grenada, Nicaragua, St Kitts-Nevis, St Vincent and the Grenadines
SOUTH AMERICA	2	French Guyana, Guyana
AUSTRALIA AND SOUTH PACIFIC	8	Kiribati, Marshall Islands, Micronesia, Nauru, Palau, Samoa, Solomon Islands, Tuvalu
	55	

Table 1.10 Countries not using UDC[173]

SURVEYED COUNTRIES (208)	**CONFIRMED NOT USING UDC IN 2004**	
ASIA	11	Burma (Myanmar), Hong Kong, Iran*, Lebanon, Nepal, Oman, Pakistan, Palestine, Phillippines*, Thailand, United Arab Emirates
AFRICA	10	Botswana, Egypt*, Ethiopia, Gambia, Malawi, Mauritius, Namibia, Sierra Leone, Zambia, Zimbabwe
NORTH AND CENTRAL AMERICA	14	Aruba, Bahamas, Belize, Cayman Islands, Dominica, Dominican R., Guadeloupe, Haiti, Honduras, Martinique, Netherland Antilles, Panama, Puerto Rico, St Lucia
AUSTRALIA AND SOUTH PACIFIC	6	New Caledonia, Niue, Papua New Guinea, Tonga, Vanuatu, Antigua & Barbuda
	41	

173 Countries marked with asterisk are those which were known to have had UDC users in the past, (Egypt, Iran, Philippines).

Table 1.11 Countries using UDC in 2004, complete data

COUNTRY (112)	USAGE BAND[174]	LANGUAGE	UDC IN NATIONAL BIBL.	VERIFIED			DATA CONFI-DENCE[175]
				E-MAIL 2004	LITERATURE	UDCC	
			EUROPE				
1. ALBANIA	A	Albanian	●	●	●		1
2. ANDORRA	A	Spanish	●	●	●	●	1
3. AUSTRIA	C	German		●	●		1
4. BELARUS	B	Russian	●	●	●		1
5. BELGIUM	AB	Fr., Dut.		●	●	●	1
6. BOSNIA & HERZ.	A	Serb.,Cro.	●	●	●		1
7. BULGARIA	AB	Bulgarian	●	●	●		1
8. CROATIA	A	Croatian	●	●	●	●	1
9. CZECH REP	A	Czech	●	●	●	●	1
10. DENMARK	B	Danish		●	●		1
11. ESTONIA	A	Estonian		●	●		1
12. FAROE ISLANDS	C	Danish		●			1
13. FINLAND	B	Fin., En.	●	●	●	●	1
14. FRANCE	B	French	●	●	●	●	1
15. GERMANY	BC	German	●		●	●	1
16. GREECE	C	English		●	●		1
17. HUNGARY	A	Hungarian	●	●	●	●	1
18. ICELAND	C	English		●	●		1
19. IRELAND	BC	English			●	●	2
20. ITALY	C	Italian		●	●	●	1
21. LATVIA	A	Latvian	●	●	●	●	1
22. LITHUANIA	A	Lithuanian	●	●	●	●	1
23. LUXEMBOURG	B	French		●	●	●	2
24. MACEDONIA	A	Macedonian	●	●	●	●	1
25. MALTA	C	English		●	●		1

[174] Symbol C* denotes that for these countries, only a very few libraries using UDC exist.. E.g. in Sudan this is only Information Science Abstracts, in Syria this is only ICDC library, Jamaica three libraries etc

[175] 1= reliable information, confirmed by number of sources; 2= fairly reliable information; 3=further investigation desirable

COUNTRY (112)	USAGE BAND[174]	LANGUAGE	UDC IN NATIONAL BIBL.	VERIFIED			DATA CONFI-DENCE[175]
				E-MAIL 2004	LITERATURE	UDCC	
26. MOLDAVIA	A	Romanian	●	●	●		1
27. NETHERLANDS	BC	Dutch		●	●	●	1
28. NORWAY	B	Norw., En.		●	●	●	1
29. POLAND	A	Polish	●		●	●	1
30. PORTUGAL	A	Portuguese	●	●	●	●	1
31. ROMANIA	A	Romanian	●	●	●	●	1
32. RUSSIAN FEDER.	B	Russian	●	●	●	●	1
33. SERBIA & MONT.	A	Serbian	●		●	●	1
34. SLOVAKIA	A	Slovakian	●	●	●	●	1
35. SLOVENIA	A	Slovenian	●	●	●	●	1
36. SPAIN	A	Spanish	●	●	●	●	1
37. SWEDEN	B	Swedish		●	●	●	1
38. SWITZERLAND	AB	Ger., Fr., En.		●	●	●	1
39. U.K.	BC	English		●	●	●	1
40. UKRAINE	A	Ukranian	●		●	●	1
ASIA							
41. ARMENIA	AB	Russian		●	●	●	1
42. AZERBAIJAN	AB	Azeri, Ru		●	●		1
43. BANGLADESH	C	English		●	●		2
44. BHUTAN	C	English		●	●		3
45. CHINA	C	Chinese		●	●		2
46. GEORGIA	A	Georgian		●	●		1
47. INDIA	B	English		●	●	●	1
48. INDONESIA	BC	En., Ind		●	●		1
49. ISRAEL	BC	Heb., En.,		●	●	●	1
50. JAPAN	B	Japanese		●	●	●	1
51. KAZAKHSTAN	B	Russian		●	●		1
52. KYRGYZSTAN	B	Ru., Kyrgyz	●	●	●		1
53. MACAO (CHINA)	AB	Port., Chin.					2
54. MALAYSIA	C*	English		●	●		2
55. SINGAPORE	C	English		●	●		3

COUNTRY (112)	USAGE BAND[174]	LANGUAGE	UDC IN NATIONAL BIBL.	VERIFIED			DATA CONFI-DENCE[175]
				E-MAIL 2004	LITERATURE	UDCC	
56. SOUTH KOREA	C	Korean		●	●		3
57. SRI LANKA	C	English		●	●	●	1
58. SYRIA	C*	English			●	●	1
59. TAIWAN	C	Chinese			●		3
60. TAJIKISTAN	C	Russian		●	●		1
61. TURKEY	C	Turkish		●	●		2
62. TURKMENISTAN	C*	Russian		●			3
63. UZBEKISTAN	B	Russian		●	●		1
64. VIETNAM	BC	Vietnamese		●	●		2
AFRICA							
1. ALGERIA	AB	French	●		●		3
2. ANGOLA	A	Portuguese		●	●		1
3. BENIN	BC	French		●	●		2
4. BURKINA FASO	AB	French		●	●		1
5. CAMEROON	C	French		●	●		3
6. CAPE VERDE	BC	Portuguese			●		3
7. CONGO DR	A	French		●	●		2
8. GHANA	B	English		●	●		2
9. GUINEA BISSAU	B	Portuguese		●	●		2
10. KENYA	C	English		●	●		2
11. LIBYA	BC	French	●		●		3
12. MADAGASCAR	BC	French	●		●		3
13. MALI	AB	French		●	●		2
14. MOROCCO	BC	French	●	●	●		2
15. MOZAMBIQUE	AB	Portuguese		●	●		3
16. NIGER	BC	French		●	●		3
17. NIGERIA	C	English		●	●		1
18. SENEGAL	AB	French	●	●	●	●	1
19. SOUTH AFRICA	B	English		●	●		1
20. SUDAN	C*	English		●	●		1
21. TANZANIA	C	English		●	●	●	1

COUNTRY (112)	USAGE BAND[174]	LANGUAGE	UDC IN NATIONAL BIBL.	VERIFIED			DATA CONFI-DENCE[175]
				E-MAIL 2004	LITERATURE	UDCC	
22. TOGO	A	French		●	●		2
23. TUNISIA	B	French	●	●	●	●	2
24. UGANDA	BC	English		●	●		2
NORTH AND CENTRAL AMERICA							
1. BARBADOS	C*	English		●	●		1
2. CANADA	B	En., Fr		●	●	●	1
3. COSTA RICA	BC	Spanish		●	●		2
4. CUBA	B	Spanish			●	●	3
5. GUATEMALA	C	Spanish		●	●	●	1
6. JAMAICA	C*	English		●	●		1
7. MEXICO	C	Spanish		●	●	●	3
8. TRINIDAD & TOB.	C	English		●	●	●	1
9. USA	C	English		●	●	●	1
SOUTH AMERICA							
1. ARGENTINA	B	Spanish	●	●	●	●	1
2. BOLIVIA	BC	Spanish		●	●		2
3. BRAZIL	B	Portuguese		●	●	●	1
4. CHILE	BC	Spanish		●	●	●	2
5. COLOMBIA	BC	Spanish		●	●		3
6. ECUADOR	BC	Spanish		●	●		3
7. PARAGUAY	C	Spanish		●	●		3
8. PERU	C	Spanish		●	●	●	1
9. SURINAM	BC	Dutch		●	●		1
10. URUGUAY	B	Spanish		●	●	●	1
11. VENEZUELA	C	Spanish			●		3
AUSTRALIA AND SOUTH PACIFIC							
1. AUSTRALIA	BC	English		●	●	●	1
2. FIJI	C	English		●	●		2
3. FRENCH POLY.	BC	French		●			1
4. NEW ZEALAND	BC	English		●	●	●	1

9 UDC TRANSLATIONS: BIBLIOGRAPHIC SURVEY

9.1 Reports on the UDC translations up to 2004

Table 1.12 Summary report on languages in Dubuc's bibliography, status in 1971 (Dubuc, 1973)

EDITIONS	NO	LANGUAGES
FULL	10	Czech[176], English, French, German, Hungarian, Japanese, Polish, Portuguese, Russian, Spanish
MEDIUM	2+7	German, Russian In Prep.: English, Macedonian, Polish, Portuguese, Spanish, Serbian[177], Slovenian
ABRIDGED	17	Dutch, German, English, Spanish, Finnish, French, Hebrew, Hungarian, Italian, Japanese, Macedonian, Polish, Portuguese, Norwegian, Serbo-Croatian, Slovenian, Swedish
SPECIAL	14	Czech, Danish, Dutch, English, French, German, Hungarian, Icelandic, Polish, Portuguese, Russian, Serbian, Spanish, Swedish Multilingual German/English/Japanese [Pharmacology]
TOTAL	21	Czech, Danish, Dutch, English, Finnish, French, German, Hebrew, Hungarian, Icelandic, Italian, Japanese, Macedonian, Norwegian, Polish, Portuguese, Russian, Serbian, Slovenian, Spanish, Swedish

Table 1.13 Rigby's list of UDC translations 1981 (Rigby, 1981)

EDITIONS	NO	LANGUAGES
FULL	12+1	Czech[178], English, French, German, Hungarian, Italian, Japanese, Polish, Portuguese, Russian, Serbian, Spanish In preparation: Romanian
MEDIUM	9+4	French, German, Italian, Japanese, Polish, Portuguese, Russian, Serbian, Slovenian In Prep.: English, Macedonian, Slovakian, Spanish
ABRIDGED	19+2	Czech, Dutch, English, Finnish, French, German, Hebrew, Hungarian, Italian, Japanese, Korean, Macedonian, Polish, Portuguese, Romanian, Serbian, Slovenian, Spanish, Swedish In Prep.: Arabic, Danish Multilingual 1958-68
SPECIAL	18	Czech, Danish, Dutch, English, French, German, Hungarian, Italian, Japanese, Norwegian, Polish, Portuguese, Romanian, Russian, Serbian, Spanish, Swedish Multilingual 1977
TOTAL	24	Arabic, Czech, Danish, Dutch, English, Finnish, French, German, Hebrew, Hungarian, Italian, Japanese, Korean, Macedonian, Norwegian, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Slovenian, Spanish, Swedish

[176] While listing a full edition with the title in the Czech language and published in Prague, Dubuc uses expression 'Czechoslovak'.
[177] Dubuc uses expression Serbo-Croatian when listing edition in Serbian.
[178] Rigby lists a full edition that is obviously in Czech as Czechoslovak. He also noted a Serbian translation as Serbo-Croat.

Table 1.14 FID bibliographical survey summary data, status in 1980 (A bibliographical survey, 1982)

EDITIONS	NO	LANGUAGES
FULL	13	Czech[179] ; Slovak, English, French, German, Hungarian, Italian, Japanese, Polish, Portuguese, Russian, Serbian , Spanish
MEDIUM	11	English, French, German, Italian, Japanese, Polish, Portuguese, Russian, Serbian, Slovak (in preparation), Slovenian
ABRIDGED	19	Czech, Dutch, English, Finnish, French, German, Hebrew, Hungarian, Italian, Japanese, Korean, Macedonian, Polish, Portuguese, Romanian, Serbian, Slovenian, Spanish, Swedish Trilingual German/French/English edition
SPECIAL	14	Danish, Dutch, English, French, German, Hungarian, Italian, Japanese, Norwegian, Polish, Portuguese, Russian, Spanish, Swedish
TOTAL	23	Czech, Danish, Dutch, English, Finnish, French, German, Hebrew, Hungarian, Italian, Japanese, Korean, Macedonian, Norwegian, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Slovenian, Spanish, Swedish

Table 1.15 Reports on translations in the E&C 1993-2003

EDITIONS	NO	LANGUAGES
FULL	4	English , Hungarian, Polish, Russian
MEDIUM [180]	11	(E)Czech, Croatian, Dutch, (E)English, French, Hungarian, (E)Japanese, Romanian Serbian, (E)Slovak, (E)Spanish,
ABRIDGED	9+1	Croatian, Czech, Danish, English, French, Polish, Romanian, Slovenian, Spanish In Prep.: German
SPECIAL		English, Ukrainian
ELECTRONIC	9	DB FILE: English (MRF); Lithuanian/Engl (MRF); Slovenian (abr.), Russian (full) CDROM: Czech/Eng (MRF), Japanese/Eng (MRF); Russian(full), Slovak/Eng (MRF); Spanish (MRF) WEB: Czech (MRF); English (MRF); Swedish (abridged)
TOTAL	19	Czech, Danish, Dutch, English, EstonianFrench, German, Hungarian, Japanese, Lithuanian, Polish, Romanian, Russian, Serbian, Slovak, Slovenian, Spanish, Swedish, Ukrainian

179 In the FID publication edition is called Czechoslovak. Strachan explained "The Czechoslovak Full Edition was so called because some sections were translated into Czech and others into Slovak. If my recollection is correct, the title pages of each section used the Czech or Slovak word for "Czechoslovak" in the title, depending on the language of that section" (Strachan, 2004).

180 Medium in this section usually means that edition is based on the UDC MRF, which is of the size of the former medium edition. Languages marked with (E) have translations in both electronic and printed medium editions. In the section electronic medium editions are repeated.

9.2 Growth of number of UDC translations 1970-2004

Table 1.16: Comparison of languages appearing in UDC translation reports 1970 - 2004

DUBUC 1970	RIGBY 1981	FID SURVEY 1982	E&C 1993-2003	SURVEY 2004
—	—	—	—	Albanian
	Arabic	—	—	—
—	—	—	—	Azerbaijan
—	—	—	—	Bulgarian
—	—	—	—	Catalan
—	—	—	—	Chinese
—	—	—	—	Croatian
Czech	Czech	Czech	Czech	Czech
Danish	Danish	Danish	Danish	Danish
Dutch	Dutch	Dutch	Dutch	Dutch
English	English	English	English	English
—	—	—	Estonian	Estonian
Finnish	Finnish	Finnish	—	Finnish
French	French	French	French	French
—	—	—	—	Georgian
German	German	German	German	German
Hebrew	Hebrew	Hebrew	—	Hebrew
Hungarian	Hungarian	Hungarian	Hungarian	Hungarian
Icelandic	—	—	—	Icelandic
—	—	—	—	Indonesian
Italian	Italian	Italian	—	Italian
Japanese	Japanese	Japanese	Japanese	Japanese
—	Korean	Korean	—	Korean
—	—	—	—	Kyrgyz
—	—	—	—	Latvian
—	—	—	Lithuanian	Lithuanian
Macedonian	Macedonian	Macedonian	—	Macedonian
Norwegian	Norwegian	Norwegian	—	Norwegian
Polish	Polish	Polish	Polish	Polish
Portuguese	Portuguese	Portuguese	—	Portuguese
—	Romanian	Romanian	Romanian	Romanian
Russian	Russian	Russian	Russian	Russian
Serbian	Serbian	Serbian	Serbian	Serbian
—	Slovakian	Slovakian	Slovakian	Slovakian
Slovenian	Slovenian	Slovenian	Slovenian	Slovenian
Spanish	Spanish	Spanish	Spanish	Spanish
Swedish	Swedish	Swedish	Swedish	Swedish
—	—	—	—	Turkish
—	—	—	Ukrainian	Ukrainian
—	—	—	—	Vietnamese
21	24	23	19	39

9.3 BIBLIOGRAPHY: LAST REPORTED UDC TRANSLATIONS[181]

1. ALBANIAN (script Roman)

REFERENCE:	Tabela e Klasifikimit Dhjetor Universal (KDU).- [abridged edition]. - Tirank : Biblioteka Kombetare, 1991.
SOURCE:	National Library of Albania (Biblioteka Kombetare)
NOTE:	In Albania UDC is used in public libraries

2. AZERBAIJANI (script Roman)

REFERENCE:	**[reference requested]**
SOURCE:	E-mail correspondence in 2004 with National Academy of Science in Baku confirmed that there was an early abridged translation of the UDC in Azerbaijani in 1960s. Upon this the reference has been requested and was not received at the time of this report
NOTE:	UDC is used in Azerbaijan in (technical, scientific libraries)

3. BULGARIAN (script Cyrillic**)**

REFERENCE:	Tablici na desiticnata klasifikacij [Schedules of the decimal classification] / Redaktor Aleksandra Nikolova Dipcikova et al. - [?medium ed.]. -Sofija : Narodna Biblioteka Sv. Kiril i Metodij, 1992.
SOURCE:	OPAC: http://nsk-libsu.uni-sofia.bg/ab/eab.HTM

4. CATALAN (script Roman)

REFERENCE:	Classificació decimal per a les biblioteques catalanes / Dirigida per Jordi Rubió i Balaguer, preparada per Jordi Rubió i Lois. 4th ed. Barcelona : Teide, [1982].
SOURCE:	OPAC: http://ccuc.cbuc.es/

5. CHINESE (script Chinese)

REFERENCE:	Guo ji shi jin fen lei fa = Classificacao decimal universal = Universal decimal classification / Aomen zhong yang tu shu guan bian yi = redaccao e traducao: Biblioteca Central de Macau. - Zhong wen ban jian ben = edicao chinesa simplicada = simplified edition Chinese text . - Macau : Aomen wen hua si shu = Instituto Cultral de Macau, 1997
SOURCE:	National Library of China, Acquisitions & Cataloging Department
NOTE:	13 bibliographic references to the UDC schedules were found in National library catalogue, many dating from the 1960s and related to special editions. Here included is the last edition mentioned.

[181] Status as of December 2004.

6. CZECH (script Roman)

REFERENCE: MDT/UDC : Mezinárodní desetinné třídění : CD ROM: English/Czech. 5th ed. Praha : Národní knihovna ČR ; Beroun : AiP, 2003.
MDT/UDC : Mezinárodní desetinné třídění. 2001-2003 http://aip.nkp.cz/mdt/

SOURCE: E&C

NOTE: There are three publications of the Czech standard edition based on the UDC MRF published on CD ROM. In 1997 the CD ROM was Czech only, while the editions in 1998, 2000 and 2003 were bilingual English/Czech. Czech UDC online was first published in 2001 and is updated regularly.

7. CROATIAN (script Roman)

REFERENCE: Univerzalna decimalna klasifikacija : hrvatsko srednje izdanje / Edited by Tomislav Murati. - Zagreb: Nacionalna i sveučilišna knjižnica. 2004 - .

Univerzalna decimalna klasifikacija. Hrvatsko džepno izdanje. Zagreb: Naklada Dominović, 2003. [Translation of the Universal Decimal Classification: pocket edition. BSI: 1999].

SOURCE: National and University Library, Zagreb Croatia, confirmed that the 1st volume of the medium edition is in print in 2004 and provided the above reference

NOTE: First translation of the UDC in Croatian appeared in 1968 (abridged ed.) it was followed by special schedules for public libraries in 1984, and special scheme for school libraries in 1991. References available in OPAC at http://www.nsk.hr/opac-crolist/uvod.html

8. DANISH (script Roman)

REFERENCE: Danskforkortet UDK: Grupperne 629 Trafikmidler, 656 Trafikorganisation (med alfabetisk register) Lyngby, Dansk Central for Dokumentation - DCD & Den Danske UDK-komite, 1995.

Dansk forkortet UDK: med alfabetisk register / Dansk Central for Dokumentation & Den danske UDK-komité. - Kopenhagen: Dansk Central for Dokumentation - DCD, 1986.

SOURCE: OPAC http://bibliotek.dk/ , correspondence with Royal School of Library and Information Science, E&C

9. DUTCH (script Roman)

REFERENCE: UDC. Tabellen. 15e dr. Houten, Bohn Stafleu Van Loghum bv, 1993. 4 Suppl. 14, 15, 16
Universele Decimale Classificatie - tabellen. Volume ?. The Hague: UDC Consortium, 2003.

Universele Decimale Classificatie - tabellen. 15e druk. Aanpassing 1999. Volume 21. Alphen aan den Rijn, Samsom, 1999

SOURCE: E&C

NOTE:	Translation of the UDC to Dutch is based on the UDC MRF and is published in supplements. The last reference published in E&C 25 2003 does not have Volume number.

10. ENGLISH (script Roman)	
REFERENCE:	UDC-online.London: BSI, Technical Indexes, 2001-2004, http://www.udc-online.com UDC Abridged Edition. London: BSI, 2003. Universal Decimal Classification. International Medium Edition. English text, Edition 2. [BS 1000M: 1993]. London, BSI Standards, London 1993.
SOURCE:	E&C
NOTE:	There were two abridged editions in English: pocket edition in 2000 and its update titled Abridged edition in 2003. UDC-online was first published in 2001 and was updated in 2002. New version will be released by BSI only in January 2005.

11. ESTONIAN (script Roman)	
REFERENCE:	UDK. Universaalne Detsimaalklassifikatsioon. Liigitustabelid / Toim kolleegium: Marje Aasmets, Arda-Maria Kirsel, Malle Lang, Sirje Nilbe, Aili Normak, Maara Üksti. [medium ed.].-Tallinn: Eesti Vabariigi Kultuuriministeerium, 1999.
SOURCE:	E&C

12. FINNISH (script Roman)	
REFERENCE:	Yleinen kymmenluokittelu (UDK) [Universal Decimal Classification (UDC)] /[publ.] Tietopalveluseura. - Abridged Finnish version, 3rd, rev. ed. - Helsinki : Tietopalveluseura [Finnish Society for Information Services], 1983. - In loose binding. - [Supplements in 1991].)
SOURCE:	National Library of Health Sciences

13. FRENCH (script Roman)	
REFERENCE:	Classification Décimale Universelle. Edition Moyenne Internationale en 3 volumes. Liège: Éditions du C. É. F. A. L. 2004. CDU Édition abrégée. 7e éd. Liège: Éditions du CÉFAL, 2001
SOURCE:	E&C, publisher's catalogue at http://www.cefal.com/

14. GEORGIAN (script Mkhedruli)	
REFERENCE:	Sabiblioteko klasifikaciis cxrilebi [Tables for Library Classification]. Ed. G. Takniashvili. - 1998, 117 pp. - Based on the 3rd Russian edition.
SOURCE:	Library Association in Georgia
NOTE:	Abridged edition, used to transform public libraries from BBK to UDC. New abridged edition in preparation.

15. GERMAN (script Roman)

REFERENCE: Dezimalklassifikation : internationale mittlere Ausgabe / herausgeber, DIN Deutsches Institut für Normung E. V. - 2. Aufl. der DK-Handausg. I serie: FID ; 550. Berlin : Beuth, 1978-1985. 2 bind.

SOURCE:

NOTE: Abridged German edition in preparation by C.É.F.A.L., publisher in Belgium was announced back in 2002.

16. HEBREW (script Hebrew)

REFERENCE: Universal decimal classification: abridged Hebrew edition. - Tel Aviv, Israel:. Center of Scientific and Technological Information, 1969

SOURCE: Elyachar Cental Library, UDC Classification Dept. Technion Israel Institute of Technology

NOTE: Reference was not made available in the language of original. Libraries in Israel are using English editions

17. HUNGARIAN (script Roman)

REFERENCE: Egytemes Tizedes Osztályozás. -Teljes kiadás 2 Vallas. -Budapest, Magyar Szabványügyi Hivatal, 1992.

SOURCE: E&C

18. ICELANDIC (script Roman)

REFERENCE: Tugstafakerfi við bréfafærslu og skrásetningu bæjar- og sveitarstjórnamálefna / [prepared and translated by] Sigurbjörnsson (Lárus). - Prentað sem handrit [printed from manuscript]. - Reykjavík: Skjalasafn Reykjavíkurbæjar, S.d [1953]

SOURCE: Dubuc, 1973, verified 2004

NOTE: The reference is verified and corrected by a librarian in IceTec - Technological Institute of Iceland, Reykjavik in 2004. Dubuc's date is changed from 1957 to 1953 according to the National Library of Iceland database http://www.gegnir.is . This is a special edition of a decimal classification system used in municipal archives, especially in Reykjavik. Information on the book states that this is an abridged edition containing only 351 principal divisions. Translation is based on the Danish edition "Dansk kommunal decimal klassifikation, Koldin, 1951" which is again translated/adapted from the German: Dezimalklassifikation. Deutsche Gesamtausgabe 1934-48.

19. INDONESIAN(Bahasa Indonesia) (script Roman)

REFERENCE: [full reference requested]

SOURCE: Dept of Library and Information Science, Universitas Indonesia
Depok, Indonesia

NOTE: Part of 636 (Forestry) has been translated into Indonesian Language circa 1967 in Bogor; 34 (Law) translated by National Legal Documentation Centre in early 1970s (all were out of print).In Indonesia, UDC is used in special libraries mainly among agricultural and forestry libraries

20. ITALIAN (script Roman)

REFERENCE: La CDU Online. - [an abridged ed. from 1987]. - http://mail.biocfarm.unibo.it/%7Espinelli/cdu/

Classificazione decimale universale : edizione abbreviata italiana / Consiglio Nazionale delle Ricerche, Istituto di studi sulla ricerca e documentazione scientifica. - Roma : Edizioni dell'Ateneo, 1987.

Classificazione decimale universale. Ed. completa italiana. Roma : Consiglio nazionale delle ricerche, Laboratorio di studi sulla ricerca e sulla documentazione, 1972-1985. 20 v. (Pubbl. FID ; 479).

SOURCE: OPAC BSB: http://www.cultura.toscana.it/biblioteche/bsb/index.shtml

21. JAPANESE (scripts hiragana/katakana/kanji)

REFERENCE: Kokusai jusshin bunruiho (UDC) [International Decimal Classification] / [compiled by] Joho Kagaku Gijutsu Kyokai [Association for Technologies of Information Science]. CD-ROMban [CD-ROM ed.]. - Tokyo: Joho Kagaku Gijutsu Kyokai, 2004.

[UDC Japanese version with Retrieval System]. CD-ROM. [UDC - standard edition, bilingual English - Japanese edition]. Tokyo: INFOSTA, 2002.

Kokusai Zyusshin Bunruihyo. Nihongo chukanban Dai 3 Han[UDC Japanese 3rd medium ed]. Tokyo: Joho Kagaku Gijutsu Kyokai [Maruzen Co.].- 1994.

SOURCE: Agriculture, Forestry and Fisheries Research Information Center;
Waseda University Scholarly Information Network System
http://wine.wul.waseda.ac.jp/screens/mainmenu.html; E&C

22. KOREAN (scripts: Hangeul, Hanja)

REFERENCE: Kukche Sipchin Pullyupop: Han'gugo kallyakp'an / Han'guk Kwahak Kisul Chongbo Sent'o p'yon. [=Universal Decimal Classification: abridged Korean edition]. Soul : Han'guk Kwahak Kisul Chongbo Sent'o, 1973. 284 p. ; 25 cm. FID ch'ulp'anmul ; che 505-ho

SOURCE: National Assembly Library of Korea (Ms Mikyung Song) and transliteration according to McCune-Reischauer romanization system.

23. KYRGIZ (scripts: Cyrillic, Roman, Arabic)

REFERENCE: Seloduk kitepkanalar uchun klassifikatsiyanyn kyskacha tablitsiyalary: oruscha 2-basylyshynan kotoruldu/RSFSRdin Madaniyat Ministerstvosu.- Frunze, 1961

SOURCE: Technologies Department The National Library of the Kyrgyz Republic

NOTE: Kyrgyzstan: 70% libraries use BBK. UDC is used mostly by special libraries in Kyrgyzstan. But from the 1980s some academic libraries began implementing UDC. Most of these libraries classify their collection with Russian UDC schedules.

24. LATVIAN (script Roman)

REFERENCE: Klasifikācijas tabulas publiskajām bibliotēkām / Sastādītāja Lilija Grīnfogele ; zinātniskās redaktores Gertrūde Kučinska un Sarmīte Morica ; terminologijas redaktore Anna Mauliņa. - [abridged] - Rīga : Latvijas Nacionālā bibliotēka, 1996.

SOURCE: OPAC, verified by National Library of Latvia

NOTE: This is a special edition for public libraries.

25. LITHUANIAN (script Roman)

REFERENCE: Universalioji dešimtainė klasifikacija : trumposios klasifikacijos lent. / Lietuvos nac. M. Mažvydo b-ka ; sudarytoja S. Šimakauskienė. - Vilnius : Lietuvos nac. M. Mažvydo b-ka, 1993.

Universalioji Desimtaine Klasifikacija. Sutrumpintos lenteles. Vilnius, 1991.

SOURCE: OPAC http://www.libis.lt:8080/en/welcome.html

NOTE: Lithuanian translation of the UDC started in 1995 based on the MRF and is produced and maintained in a Lithuanian/English UDC database in National library of Lithuania which is used as tool for indexing and retrieval - http://www.lnb.lt. (Gobyté, 2000)

26. MACEDONIAN (script Cyrillic)

REFERENCE: Univerzalna decimalna klasifikacija. [1st. Macedonian medium ed.]. Skopje: Narodna univerzitetska biblioteka "Kliment Ohridski", 1986

SOURCE: http://nubsk1.nubsk.edu.mk/cobiss/

NOTE: script of the edition is cyrillic

27. NORWEGIAN (script Roman)

REFERENCE: Universelle desimalklassifikasjon : oppdragelse, undervisning, fri tidsbeskjeftigelse ; arbeidskopi oversatt etter "Universal decimal classification special edition for education" FID.374/1965". Trykt: Oslo, 1965.

SOURCE: OPAC: http://www.nb.no/baser/bibliotek/indexnew3.html

28. POLISH (script Roman)

REFERENCE: Uniwersalna Klasyfikacja Dziesietna: publikacja nr UDC-P022 autoryzowana przez Konsorcjum UKD nr licencji UDC-9709. Wydanie skrócone dla biezacej bibliografii narodowej i bibliotek publicznych. Warsaw: Biblioteka Narodowa, 1997

Uniwersalna Klasifikacja Dziestetna. II wydanie pelne w jezyku polskim. Tablica lc. Dzial 8. Warsaw, Instytut Informacji Naukowej, Technicznej i Ekonomicznej, 1996

SOURCE: E&C

29. PORTUGUESE (script Roman)

REFERENCE: Classificação Decimal Universal. Edição-padrão internacional em língua portuguesa. Brasília : IBICT, 1997 - 1999. Parte 1 Tabelas sistemáticas. 1997. Parte 2 Índice alfabético. 1999.

SOURCE: http://www.ced.ufsc.br/~ursula/5213/

30. ROMANIAN (script Roman)

REFERENCE: The Universal Decimal Classification: tables and alphabetical index. Abridged edition (Classificare Zecimala Universala: tabele si index. Editie prescurtata); editie ingrijita de A. Andrian, G. Clinca. - Bucharest, ABBPR, 1995.

SOURCE: E&C

31. RUSSIAN (script: Cyrillic)

REFERENCE: Univerzaln'aja deseticnaja klassifikacija (UDK): izm. i dop. [corrections and extensions]. Prepared by T.S. Astahova; editor in chief Ju. M. Arskij. Moskva : VINITI, 2003. Vjip 2

Univerzaln'aja deseticnaja klassifikacija (UDK): T 4: 55/59 Geologiceskije i biologiceskije nauki [Geological and biological sciences]. Prepared by Ruska Akademija Nauk , VINITI; editor in chief Ju. M. Arskij. Moskva : VINITI, 2003

Univerzaln'aja deseticnaja klassifikacija (UDK): elektronn'iij resurs [electronic form]: baza dann'ih. Prepared by Gos. public. nauc.-tehn. biblioteka. Moskva : GPNTB, 2001.

Universal'naja decimal'naja klasifikacija v II tomah; obsee redaktirovanie V. I. Fedosimov. 4-e polnoe izdanie.(Publication no UDC-P16/2, authorized by UDC Consortium). Moskva, Rektor, 1999

SOURCE: E&C + OPACs

NOTE: Russian full edition based on the German full edition is completed and stored in the database in 1994. This is used as a base for publication of full editions (Arskiy & Nesterov, 2000)

32. SERBIAN (script Cyrillic, occasionally Roman)

REFERENCE: Univerzalna decimalna klasifikacija. Deo 1,[part 1: sistematske tablice [systmatic tables] / editor S. Simonovic-Mandic; translators S. Dukanovic [et al]. 2nd Serbian medium ed. Beograd : YUBIN, 2004

Univerzalna decimalna klasifikacija : izmene i dopune [extensions and corrections] / editor in chief Svetlana Simonovic-Mandic. Beograd : Beograd : Jugoslovensko bibliografsko-informacijski institut YUBIN, 2003.

SOURCE: OPAC: http://vbs.nbs.bg.ac.yu/cobiss/

NOTE: Serbian translations of the UDC are published in Roman script

33. SLOVAKIAN (script Roman)

REFERENCE:	Medzinárodné desatinné triedenie (SK/EN) : sieťová licencia .- 2000. - (CD-ROM) [1st Slovak-English version issued on CC-ROM]
SOURCE:	E&C

34. SLOVENIAN (script Roman)

REFERENCE:	Univerzalna Decimalna Klasifikacija: abecedni register / [contributors Vili Kogovšek, Vekoslava Ludvik, Marko Švajger]. - 2nd Slovenian abridged ed., changed and extended. - Ljubljana: Centralna tehnčka knjižnica, 1994. [alphabetical index to the changed and improved abridged edition from 1983]
SOURCE:	OPAC http://www.cobiss.si
NOTE:	Slovenian translation of class religion and theology (medium edition) was published in 1973 in Belgrade (translation from German).

35. SPANISH (script Roman)

REFERENCE:	Clasificación Decimal Universal - CD-ROM. Madrid, AENOR, 2000. Clasificatión Decimal Universal. - [7th abridged ed.]- Madrid, AENOR, 2001. Clasificación Decimal Universal - Impresa.Volumen 1 - Tablas sistemáticas (0 / 5), Volumen 2 - Tablas sistemáticas (6 / 9), Volumen 3 - Indice alfabético. Madrid, AENOR, 2000.
SOURCE:	E&C

36. SWEDISH (script Roman)

REFERENCE:	Universella Decimalklassifikationen. Svensk Elektronisk Utgåva [UDC Swedish electronic abridged ed.] / redaktör Benito Miguel. - Borås: Bibliotekshögskolan, 2003. http://www.hb.se/bhs/udk/ Universella decimalklassifikationen : svensk fullständig upplaga : UDK 331 Arbete, arbetsvetenskap [UDK Class 331]. Trykt: Stockholm : Tekniska litteratursällskapet, 1983. Sidetall: 45 bl. Universella decimalklassifikationen, svensk förkortad upplaga : revideringar 1977-81. Stockholm : Tekniska litteratursällskapet, 1981 [with supplements published 1977-1994]
SOURCE:	OPAC: http://www.opac.ssb.stockholm.se
NOTE:	Swedish electronic abridged edition exists since 2001 and was last updated in 2003.

37. TURKISH (script Roman)	
REFERENCE:	Evrensel Onlu Siniflama sistemi / translated by Ceviren Fehmi Ethem Karatay. Istanbul : [not published], 1925 [+ extensions later].
SOURCE:	Robert College Library, Istanbul, Turkey
NOTE:	This translation is available at Istanbul University Central Library

38. UKRAINIAN (script Cyrillic)	
REFERENCE:	Universal'na desjatkova klasifikacija (UDK): 5 Matematika ta prirodnici nauki. Per. z angl; golov. red. M. I. Sencenko.(Publication no UDC-P033, authorized by the UDC Consortium). Kijev, Knizkova palata Ukraini, 1999
SOURCE:	E & C

39. VIETNAMESE (script Quốc Ngữ)	
REFERENCE:	Khung phan loai thap phan bach khoa: Dich tu ban rut gon tieng Nga [Universal Decimal Classification: Translated from Russian abridged edition] / Nguoi dich [translated by]: Ta Ba Hung va nhung nguoi khac; Bien tap : Vu Dinh Tuan, Nguyen Khac Bao, Truong Cam Bao . - Ha Noi: Vien Thong tin KHKT TU [Central Institute for Scientific and Technical Information], 1984. [Khoang 400 tr.]
SOURCE:	Reference provided by the Editor of Vietnamese DDC14 (Mr. Vu Van Son)
NOTE:	UDC is used in technical and scientific libraries

10. UDC DATA

10.1 VALIDATION OF UDC MRF 2003[182]

Table 1.17 Examples of variants of style and structure in text fields introduced in 1993-2004

CLASS PRESENTATION	REVISED/ INTRODUCED
ENGLISH NAME WITH/WITHOUT ORIGINAL COUNTRY NAMES	
(611) Tunisia. Republic of Tunisia. (630) Ethiopia. Federal Democratic Republic of Ethiopia (Ityop'iya. Ityop'iya Federalawi Demokrasiyawi Ripeblik) (635) Eritrea. Republic of Eritrea (64) Morocco. Kingdom of Morocco (65) Algeria. Democratic and Popular Republic of Algeria (Al Jaza'ir. Al Jumhuriyah al Jaza'iriyah ad Dimuqratiyah ash Sha'biyah)	2004
(437.3)Czech Republic. Czechia (Ceská Republika. Cesko)	2001
(474.2)Estonia. Republic of Estonia	1998
(430)Germany. Federal Republic of Germany (Bundesrepublik Deutschland)	1997
(435.9)Luxembourg. Grand Duchy of Luxembourg	1993
INCONSISTENCY OF TEXT OF INSTRUCTIONS	
INSTRUCTION FOR USE OF COLON	
Use this number only if it is clear that a chronological aspect is intended. If in doubt, use colon combination with 327.57	2003
Details by :004..., :681.612/.613	2002
For specific prohibitions use colon combination e.g. 2-428.4:002 prohibited books	2000
Synthesize from common auxiliaries, special auxiliaries, and colon combination	2000
Details by colon combination with (2...)	1993
Denote the invention or discovery by :...	1993
INSTRUCTION FOR USE OF ALPHABETICAL EXTENSIONS	
Specify by A/Z	1997
Specify by alphabetic extension A/Z (Table 1h)	2000
Denote movements by alphabetic extension (Table 1h)	1998
Denote the body by (4/9) and/or alphabetic extension	1997
Subdivide alphabetically by name of genus and species of bacterium. See Table 1h	1999
DIFFERENT STYLE IN USE OF A/Z EXTENSIONS (a word, several words, abbreviations, three letters)	
632.76Leptinotarsa decemlineata	1997
579.852.11B.anth	2001
730Rodin	1998
233-784VHP [for Visva Hindu Parisad] 26-784BNA [for Bnai Brith] 27-784ROS [for Rosicrucians]	2000
791.237.6NEW [for New York]	1999

[182] Released in 2004 and valid until 2005

CLASS PRESENTATION	REVISED/ INTRODUCED
821.111SHAK [for Shakespeare] 821.133.1-2MOL [for Molière]	1993
339(436.1)HfW [for Hochschule für Welthandel]	2001
355(441.3)StC [for École Spéciale Militaire de St Cyr]	2001
USE OF BRACKETS	
UDC-#[001]: =112.24 DESCR[100]: Middle German SCOPE[110]: (Not to be confused with =122.2`04 Middle High German) INDEX[951] :=122.2; `04	1993
DECLARING SPECIAL AUXILIARY TABLE (with range, with three dots, with facet indicator only)	
338.48-1/-6Characteristics of tourism 355.69`1/`9Technical questions concerning military transport 82-1/-9Literary forms. Genres 004.3`1no record introducing special auxiliary table, table starts with the first special auxiliary 78.01 no record introducing special auxiliary table, table starts with the first special auxiliary 2-1no record introducing special auxiliary table, table starts with the first special auxiliary	1999 1993 1993 1995 2000 2000
THE USE OF THREE DOTS IN THE MAIN UDC HEADINGS	
ANNOUNCING BEGINNING OF SOME OF THE COMMON AUXILIARY TABLES =...Table 1c - Common auxiliaries of language (0...)Table 1d. COMMON AUXILIARIES OF FORM "..."Table 1g. COMMON AUXILIARIES OF TIME -0...Table 1k. COMMON AUXILIARIES OF GENERAL CHARACTERISTICS (=...)Table 1f - Common auxiliaries of race, ethnic grouping and nationality	 2002R 2001R 2003R 1993 2002R
IN PRESENTING SOME OF SPECIAL AUXILIARY NUMBERS: =...`0Origins and periods of language. Phases of development =...`282Dialects. Local and regional language. Variants and vernaculars (1=...)Place with reference to race. Ethnic zones (1:...)One place in relation to another	 1993 " " " "
IN INTRODUCING SOME SPECIAL AUXILIARY TABLES: 34.0... (0.0...)Physical features, production and use characteristics, supplementary matter, etc.	 1998 1993
IN ONE OF PRE-SYNTHESISED NUMBERS IN THE MAIN SCHEDULES: 27-9«.../1054»Church history to the Great Schism	
IN PRESENTING THE ENTIRE CLASS WITH MAIN AND SPECIAL AUXILIARY TABLES: 82...A/ZWorks of specific authors 82...A/Z1/7Kinds of edition 82...A/Z1/7].01In original language	1993 2001
DIFFERENT WAYS OF INTRODUCING SPECIAL AUXILIARY TABLE	
338.48-1/-6Characteristics of tourism 355.69`1/`9Technical questions concerning military transport 82-1/-9Literary forms. Genres 004.3`1[no record introducing special auxiliary table, table starts with the first special auxiliary] 78.01 [no record introducing special auxiliary table, table starts with the first special auxiliary] 2-1[no record introducing special auxiliary table, table starts with the first special auxiliary]	1999 1993 1993 1995 2000 2000

Table 1.18 Non-existing notation found in the field of Parallel divisions instruction - source notation (011^a)[183]

UDC NUMBER CONTAINING THE PARALLEL DIVISION INSTRUCTION	NOTATION THAT APPEARS IN THE INSTRUCTION	NOTATION ELEMENT THAT DOES NOT EXIST
002.6	06.01/.08	06.08
341.12	06.01/.09	06.09
«69»	21/26+28/29	notation should be repeated as 21/26 and 28/29 and not connected with + (plus)
338.45.02	338.246.02	338.246.02
355.236	355.33	355.33
378.141	371.21/.215	371.215
-032.43/.49	553.43/.49	553.49
622.343/.349	553.43/.49	553.49
674.032.3/.4	582.3/.4	582.3
674.031.5/.9	582.5/.9	582.5
677.146/.149	582.6/.9	582.6
677.156/.159	582.6/.9	582.6
611.068	591.4.068	special auxiliary 591.4.068 does not exist
636.068	591.4.068	special auxiliary 591.4.068 does not exist
616-002.9	593/599	593
639.24	599.5/.7	599.7
355.69`1/`9	656.01/.09	656.09
542.9	^a66.09%^a66-9	the field contain two source notations
691.6	666.1/.2	666.2
671.411/.418	669.1/.8	669.8
691.7	669.1/.8	669.8
669.054.3	669.2/.8	669.8
=030	=030.1/.9 divided as =1/=9	=030.1/.9 does not exist, it should be introduced and 011 should point to =1/=9
811	=1/=9	=1
821	=1/=9	=1
25 records containing mistake, 19 wrong notations		

Table 1.19 Non-existing notation found in scope notes (field 110)

UDC NUMBER WHERE THE NOTE APPEARS	APPEARS IN SCOPE NOTE AS	ELEMENT THAT DOES NOT EXIST
-027.531	-025.1	-025.1

[183] If a number can be subdivided by parallel division this field contains the elements which make up the parallel division: source notation and target notation. The source notation must be an existing UDC number.

UDC NUMBER WHERE THE NOTE APPEARS	APPEARS IN SCOPE NOTE AS	ELEMENT THAT DOES NOT EXIST
24-253	242-24/-257	special auxiliary 2-257 does not exist
233-286.7	233.2-242	special auxiliary 2-242 does not exist
271.2-726.54	271.2-726.52	special auxiliary 2-726.52 does not exist
271.2-726.52	271.2-726.54	special auxiliary 2-726.54 does not exist
27-789	272-789.5	special auxiliary 2-789.5 does not exist
2-284	27-284.549	special number 2-284.549 does not exist
364.62	364.08	364.08
546-4	543-4	special auxiliary number 543-4 or 54-4 does not exist
602.6	606.4	606.4
616.895.4	616.89-008.441.1	special auxiliary 616-008.441.1 or 61-008.441.1 does not exist
-055.3	616.89-008.442	special auxiliary 616-008.442 or 61-008.442 does not exist
629.4.015	625.1.03	625.1.03
26 records containing mistakes, 28 incorrect notations		

Table 1.20 Non existing notation found in applications notes (UDC MRF field 111)

UDC NUMBER WHERE THE NOTE APPEARS	APPEARS IN APPLICATION NOTE AS	ELEMENT THAT DOES NOT EXIST
+, /	592+593	593
"67"	"61/63":	"61"
"63"	(292...)	(292)
-03	-036.674	-036.674
338.46	338.246.02	there is no special auxiliary number .02 that is listed 33, 338, 338.2, 338.42 or 338.246
338.47	338.246.02	there is no special auxiliary number .02 that is listed in 33, 338, 338.2, 338.42 or 338.246
661.8`027.1	546.185	546.185
546	546.623'32'226	546.623
551.4	551.4.0...	551.4.0
551.46.0	551.4.0...	551.4.0
582.09	582.2/.9	582.2
616-008.3/.5	616-008.3	special auxiliary number 616-008.3 does not exist
616-008.3/.5	616-008.4	special auxiliary number 616-008.4 does not exist
616-008.3/.5	616-008.5	special auxiliary number 616-008.5 does not exist
669.05	669.2/.8	669.8 does not exist
669.055	669.2/.86	669.8 does not exist

UDC NUMBER WHERE THE NOTE APPEARS	APPEARS IN APPLICATION NOTE AS	ELEMENT THAT DOES NOT EXIST
669.055	669.85/.86].055.018	669.85
78.09	7.09...	special auxiliary number 7.09 does not exist
792.092	7.09...	special auxiliary number 7.09 does not exist
794.09	7.09...	special auxiliary number 7.09 does not exist
81`0	811`03, `05, `07, `09	Special auxiliary number 81'03 or 811'03 do not exist
81`0	811`03, `05, `07, `09	Special auxiliary number 81'05 or 811'05 do not exist
81`0	811`03, `05, `07, `09	Special auxiliary number 81'07 or 811'07 do not exist
81`0	811`03, `05, `07, `09	Special auxiliary number 81'09 or 811'09 do not exist
18 records containing mistakes, 21 incorrect notations		

Table 1.21 Non-existing notation found in examples of combination (UDC MRF field 115)

COMBINATION BY DIRECT ADDITION (SUBFIELD 111^A)		
UDC NUMBER WHERE THE EXAMPLE APPEARS	APPEARS IN EXAMPLE AS:	UDC NOTATION ELEMENT THAT DOES NOT EXIST
005.745	(060.533)	(060.533)
338.48-44	(202)	(202)
502.51	(282.02)	(282.02)
314.116	-026.241	-026.241
314.116	-026.243	-026.243
7.023.1	-032.548	-032.548
666.96	.022.32-032.7	-032.7
7.023.1	-034.21	-034.21
7.023.1	-034.22	-034.22
7.023.1	-034.35`6	-034.35
7.023.1	-035.437	-035.437
687.3	-037.21	-037.21
687.3	-037.21-037.31	-037.21
7.023.1	-037.21	-037.21
687.3	-037.21-037.31	-037.31
687.3	-037.31	-037.31
687.3	-037.37	-037.37
616.8-009.12	-057:003.7	003.7
364.612	(1-21)-787.6:061.237	061.237
591.5	081.15:654.937	081.15
26-535.8	#136	#136 [all non-authorised UDC numbers should be introduced with *]

COMBINATION BY DIRECT ADDITION (SUBFIELD 111^A)		
UDC NUMBER WHERE THE EXAMPLE APPEARS	APPEARS IN EXAMPLE AS:	UDC NOTATION ELEMENT THAT DOES NOT EXIST
26-535.8	#113-118 Hallel	there is no special auxiliary number 2-118
364.692	-787.9:641:613.391	613.391
364.622	-785.4:702	702
364.614	787.8:316.732	787.8
781.7	(54):789.5	789.5
364.614	878.8:37.035	878.8
364.3	-642-62«311»	«311»
364.32	-642:331.215«311»	«311»
628.171	«313»	«313»
56	"61/62"	«61»
47 records contain mistakes, 55 incorrect notations		
COLON COMBINATIONS (FIELD 115^B)		
UDC NUMBER WHERE THE EXAMPLE APPEARS	APPEARS IN EXAMPLE AS:	UDC NOTATION ELEMENT THAT DOES NOT EXIST
352	351.88(410.188COV:430.24DRE)	(430.24)
005.912	004.733.2	004.733.2
005.921.1	005.515	005.515
005.953	005.841:343	005.841
005.961	005.914.3	005.914.3
371.12	159-051	159
316.423.6	316.343.26	316.343.26
316.663	316.343.652	316.343.652
364.614	316.647.832	316.647.832
791.43	316.755.4:314.04	316.755.4
791.233	323.24	323.24
658.8	336.546.2	336.546.2
338.486.1	339.189.44	339.189.44
364.652	34...	34...
629.361	354.74	354.74
347.793	355.054.1	355.054.1
347.793	355.055	special auxiliary number 35.055 or 355.055 do not exist
347.793	355.058:61-051	special auxiliary number 35.058 or 355.058 do not exist
656.072.45	355.33	355.33
544.526	542.917	542.917
544.526	542.942/.943	542.942
544.526	542.942	542.942

COMBINATION BY DIRECT ADDITION (SUBFIELD 111^A)		
UDC NUMBER WHERE THE EXAMPLE APPEARS	APPEARS IN EXAMPLE AS:	UDC NOTATION ELEMENT THAT DOES NOT EXIST
544.526	542.942/.943	542.942/.943
544.653	542.944.13	542.944.13
544.653	542.944.14	542.944.14
544.526	542.949.41	542.949.41
544.526	542.952.1	542.952.1
544.526	542.952.6	542.952.6
544.526	542.952.614.2	542.952.6
544.556	542.952.6	542.952.6
544.653	542.952.614.2	542.952.6
544.526	542.952.614.2	542.952.614.2
544.653	542.952.614.2	542.952.614.2
544.526	542.953	542.953
364.692	613.391	613.391
791.235	615.932	615.932
356.3	619-051	619
725.1	619	619
356.3	619-051	619
725.1	621.563	621.563
622.61	622.341.1	622.341.1
338.48-53	635.115.6(569.4)	635.115.6
664.951.32	639.211	639.211
664.951.53	639.213	639.213
664.951.2	639.222.2	639.222.2
664.951.31	639.222.2	639.222.2
664.951.51	639.222.2	639.222.2
664.951.6	639.222.5	639.222.5
664.951.4	639.223.2	639.223.2
725.1	656.151.4	656.151.4
604.2	661.181.1	661.181.1
691.328.43	691.618.92	691.618.92
785.6	787.1	787.1
629.5.066.34	929.4	929.4
52 records with mistakes in notation, 58 incorrect notations		
EXAMPLES OF COMBINATIONS WITH COMPLETE NOTATION (115^C)		
UDC NUMBER WHERE THE EXAMPLE APPEARS	APPEARS IN EXAMPLE AS:	UDC NOTATION ELEMENT THAT DOES NOT EXIST
-023.3	-023.3#8	#8 [all non-authorised UDC numbers should be introduced with *]

COMBINATION BY DIRECT ADDITION (SUBFIELD 111^A)		
UDC NUMBER WHERE THE EXAMPLE APPEARS	APPEARS IN EXAMPLE AS:	UDC NOTATION ELEMENT THAT DOES NOT EXIST
616-071.2	616-056.44-071.2	there is no special auxiliary 61-056.44 or 616-056.44. Also there is no a common auxiliary number -056.44
24-31-1	24-31-144.32	there is no special auxiliary number 2-144.32
24-31-1	24-31-144.34	there is no special auxiliary number 2-144.34
24-31-1	24-31-144.36	there is no special auxiliary number 2-144.36
26«652»-86E-24	26«652»-86E-24 1QM 26«652»-86E-24 4Q175 26«652»-86E-24 4Q186 26«652»-86E-24 4Q258 26«652»-86E-24 4Q434	All non-authorised UDC numbers should be introduced with *
271.2-564.11/.43	271.2-564.11-565.8	there is no special auxiliary number 2-564.11
271.2-564.11/.43	271.2-564.12-565.9-312.8	there is no special auxiliary number 2-564.12
271.2-564.11/.43	271.2-564.15-565.2	there is no special auxiliary number 2-564.15
271.2-564.11/.43	271.2-564.17	there is no special auxiliary number 2-564.17
271.2-564.11/.43	271.2-564.19-565.3	there is no special auxiliary number 2-564.19
271.2-564.11/.43	271.2-564.21-565.35	there is no special auxiliary number 2-564.21
271.2-564.11/.43	271.2-564.23-565.36	there is no special auxiliary number 2-564.23
271.2-564.11/.43	271.2-564.25-565.4	there is no special auxiliary number 2-564.25
271.2-564.11/.43	271.2-564.27-564.17	there is no special auxiliary number 2-564.27
271.2-564.11/.43	271.2-564.29-565.44	there is no special auxiliary number 2-564.29
271.2-564.11/.43	271.2-564.31-565.72	there is no special auxiliary number 2-564.31
271.2-564.11/.43	271.2-564.33-565.73	there is no special auxiliary number 2-564.33
271.2-564.11/.43	271.2-564.35-565.79PET	there is no special auxiliary number 2-564.35
271.2-564.11/.43	271.2-564.37-565.79-442.47	there is no special auxiliary number 2-564.37
271.2-564.11/.43	271.2-564.39-565.8ASS-442.47	there is no special auxiliary number 2-564.39
271.2-564.11/.43	271.2-564.41-565.9TRA	there is no special auxiliary number 2-564.41
271.2-564.11/.43	271.2-564.43-565.8ASS	there is no special auxiliary number 2-564.43
271.2-564.11/.43	271.2-564.15-565.2	there is no special auxiliary number 2-565.2
271.2-564.11/.43	271.2-564.19-565.3	there is no special auxiliary number 2-

COMBINATION BY DIRECT ADDITION (SUBFIELD 111^A)		
UDC NUMBER WHERE THE EXAMPLE APPEARS	APPEARS IN EXAMPLE AS:	UDC NOTATION ELEMENT THAT DOES NOT EXIST
		565.3
271.2-564.11/.43	271.2-564.21-565.35	there is no special auxiliary number 2-565.35
271.2-564.11/.43	271.2-564.23-565.36	there is no special auxiliary number 2-565.36
271.2-564.11/.43	271.2-564.25-565.4	there is no special auxiliary number 2-565.4
271.2-564.11/.43	271.2-564.29-565.44	there is no special auxiliary number 2-565.44
271.2-564.11/.43	271.2-564.31-565.72	there is no special auxiliary number 2-565.72
271.2-564.11/.43	271.2-564.33-565.73	there is no special auxiliary number 22-565.73
271.2-564.11/.43	271.2-564.35-565.79PET	there is no special auxiliary number 2-565.79
271.2-564.11/.43	271.2-564.37-565.79-442.47	there is no special auxiliary number 2-565.79
271.2-564.11/.43	271.2-564.39-565.8ASS-442.47	there is no special auxiliary number 2-565.8
271.2-564.11/.43	271.2-564.12-565.9-312.8	there is no special auxiliary number 2-565.9
-029	-029:248.2	248.2
314.045	314.7.045	314.7
544.332	542.942.5:544.332	542.942.5
544.332	544.332:542.942.5	542.942.5
544.31	542.948:544.323.2	542.948
616-089.843	616.5-089.843:611.976-032:611.982	611.982
612.187	612.187.35	612.187.35
62-135	62-135.2	62-135.2
62-135	62-135.4	62-135.4
632.76	632.768	632.768
7.035.36/.37	7.035.36(430)	7.035.36
7.035.36/.37	7.035.37(430)	7.035.37
78.087.1	787.1.087.1	787.1
19 records with mistakes in notation, 48 incorrect notations		
TOTAL IN EXAMPLES OF COMBINATIONS: 118 records with incorrect notation, 161 incorrect notations found		

Table 1.22 Non-existing notation found in references (UDC MRF field 125)[184]

UDC RECORD	MISTAKE IN REFERENCE	TYPE OF MISTAKE
(735.525.2/.8)	(735.3	EDITING MISTAKES
004.43	004.655^a811.93	
004.356.2	681.625.9...For kinds of printer	
669.056.9	esp.^a667.636.2	
(734.711)	734.912	
(292.84)	(829.9)	RECORDS DO NOT EXIST
(99)	(829.9)	
"19"	"312"	
"2"	"313"	
387	314.3	
387	314.42	
387	314.5	
392.532	314.57	
005.642.8	331.16	
342.726	342.83	
329.75	355.292	
356.21	355.333	
654	355.343.1	
357.56	355.343.2	
341.33	356.33	
355.72	356.33	
355.353.5	359.34	
343.815	376.58	
371.543.72	376.58	
631.117.2	378.663	
371.15	396	
785.12	788	
616-089.818.6	615.844.6	
338.48	910.4^a379.85	
33 records contain mistakes		

[184] The total amount of invalid notations in the field 125 is 394. Apart from the invalid notations listed here there are also 104 records that contain some kind of pre-synthesised notation in the references that cannot be linked to any record in UDC MRF and should be split so the function of linking can be established. These are not listed here because of their large number.

10.2 UDC macrostructure and notation synthesis

UDC macrostructure, as shown in the Figure 10.1 contains three kinds of UDC numbers organized into three kinds of tables: **main table** (disciplines/subdisciplines), **special auxiliary tables** (used to specify concepts within one (sub)discipline) and **common auxiliary tables** (containing generally applicable concepts).

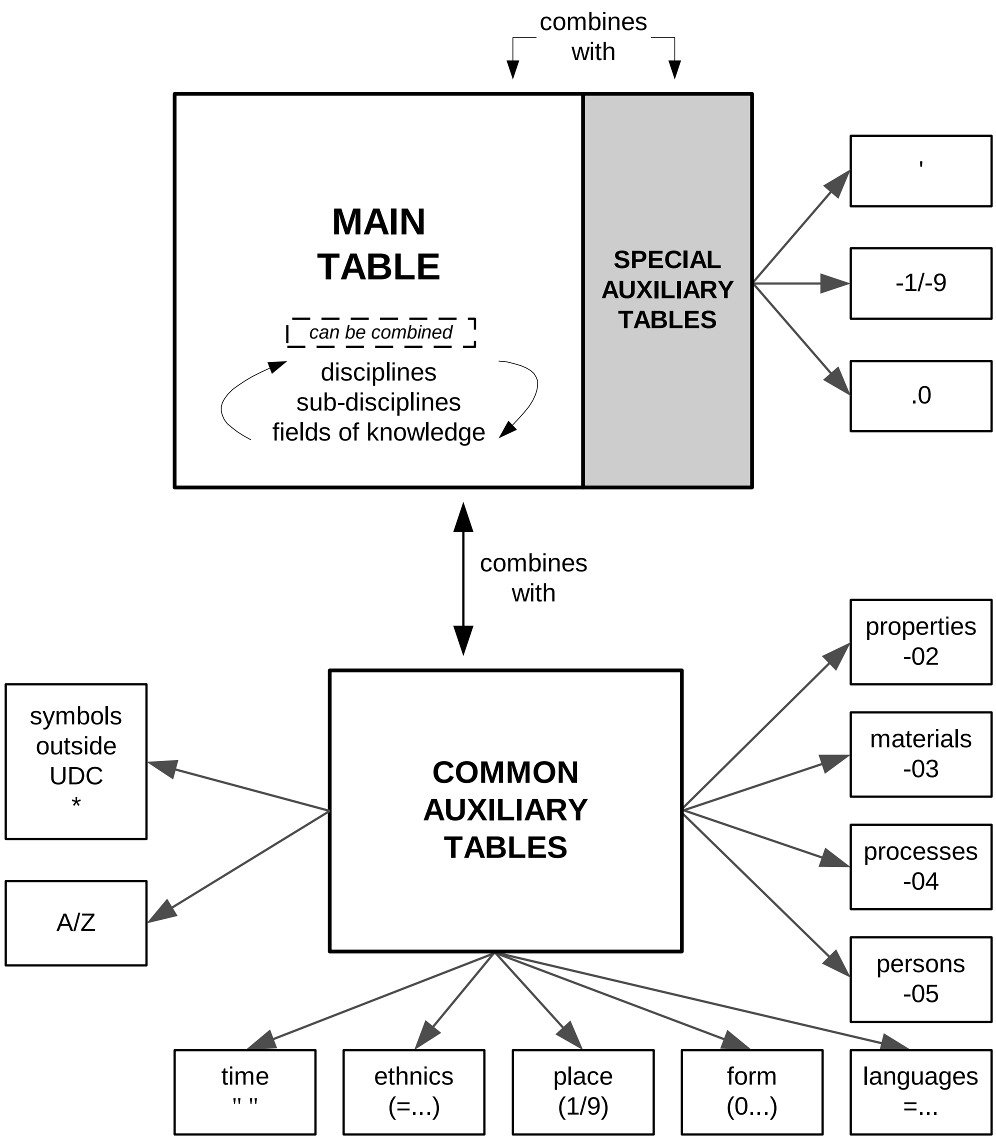


Figure 1.13 UDC macrostructure

Main table, or main schedule, consists of nine broad discipline areas (54,078 main UDC numbers [185]. Main table also 'nests' a large number of special auxiliary tables. Each (sub)discipline is likely to have at least one special auxiliary table containing special auxiliary numbers that can be combined with any main number within the same class.

Special auxiliary tables are usually used to express different broad facet categories of the same subject: such as kinds, parts, processes, materials etc., of the field of knowledge. There are 16,921 special auxiliary numbers in special auxiliary tables (out of 54,078 numbers in the main table).

In the common auxiliary tables there are nine facets of generally applicable concepts (containing 12,659 numbers) that can be combined with numbers from main table or with one another:

- Common auxiliaries of **language** (1,365)
- Common auxiliaries of **form** (360)
- Common auxiliaries of **place** (9,054)
- Common auxiliaries of **ethnic grouping and nationality** (38 [186])
- Common auxiliaries of **time** (284)
- Common auxiliaries of general characteristics:

properties (800)

materials (152)

relations, processes and operations (333)

persons and personal characteristics (267)

Common auxiliary tables can also 'nest' special auxiliary tables.

In general, synthesis in UDC is based on the fact that notations coming from different tables have different beginnings (i.e. a facet indicator), so when two or more notations are put together it is possible to determine where one element finishes

[185] Statistics in this section is based on the MRF 2004 database which released in March 2005.

[186] Ethnic grouping is derived from common auxiliaries of languages. Each notation for language can be used to build a notation for ethnic grouping.

and another begins[187]. When this is not the case there are other solutions are introduced.

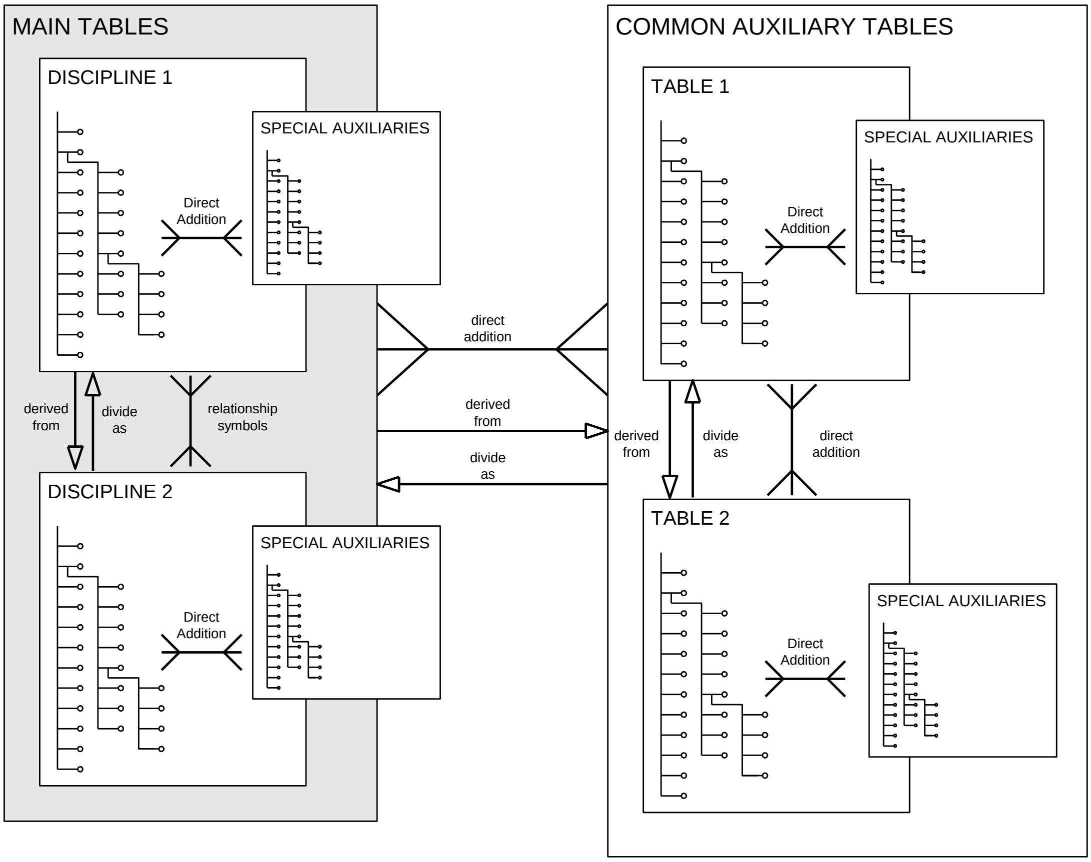


Figure 1.14 UDC synthesis diagram

It is possible to make distinction between three kinds of synthesis in UDC (represented in the Figure 10.2 with four kinds of connectors):

***i)* direct addition of one notation to another**

This rule is valid in all cases where facet indicators allow two notations to be connected while preserving the integrity of each element. Hence this synthesis is possible for the whole system between:

- any notation from main table and one or more special auxiliaries

821.111-1

[187] While main numbers have no facet indicator, special auxiliairy numbers always begin either with -1/-9 (hyphen), .0 (point nought) or ' (apostrophe). Common auxiliary always have one of the following forms: *=...*, *(0..)*, *(1/9)*, *(=1/9)*, *"..."* or *-02*, *-03*, *-04*, *-05*.

271-533-526.62

- any notation from a main table and one or more common auxiliaries

51(410)

- two or more notations from common auxiliaries

"18"(410)

- between common auxiliaries and their special auxiliary numbers

(410-18)

***ii)* combination with relationship symbols**

This rule is valid whenever two joined notations would blend resulting in a loss of distinction between the elements. Symbols (/, :, ::, +, []) are then used to separate the elements:

- two main numbers put together

32:33

- two common auxiliaries from the same table put together

(44:45)

***iii)* a notation is derived from another notation** (parallel derivation and parallel subdivision)[188]

The characteristic of this synthesis is that two notational elements blend into new notation and separate elements are not easily distinguishable. This kind of synthesis can be based on a parallel derivation rule which is stated for such a situation and prescribes the way in which the notation is created. The source notation is changed and is added to a prescribed base number to create a 'target' notation.

=111 English language

(=111) English (people)

811.111 Linguistics - English language

821.111 English literature

Also a notation can be further subdivided by adding a part of another number from elsewhere (parallel division):

Common auxiliary number *-036.5Plastics in general, especially semi synthetics* can be further subdivided using subdivision of *678.5 Plastics*, like so:

- *036.54* (based on *678.5***4** *Plastics based on cellulose or its derivatives*)

- *036.542* (based on 678.5**42** *Plastics based on cellulose or regenerated cellulose)*

- *036.542.3 (based on* 678.5**42.3** Regenerated cellulose)

The first two kind of notation synthesis (i and ii respectively) are applicable at the whole system level. Parallel derivation/division is possible only in prescribed cases and is not considered very popular among users or desirable by the UDC schedule developers who are trying to find a way of replacing this kind of synthesis with solutions described in i) and ii).

[188] This kind of synthesis is typical for number building in DDC and represents a smaller part of number synthesis in UDC.

10.3 UDC Data element schema

	ELEMENT NAME	DEFINITION	VALUE TYPE[189]	MANDATORY/OPTIONAL	REPEATABLE/NON-REPEATABLE
1	**UDC HEADING**	Single or pre-synthesised UDC notation contained in its own record	NOTATION	M	N
1.1	UDC HEADING ID	Automatically generated, unique identifier of any UDC classmark introduces since the database was created (1993)	NUMERIC	M	N
1.2	UDC HEADING TYPE	Type of notation expressed with coded value for the purpose of management and control of pre-synthesised numbers List of codes C single common auxiliary CS common auxiliary with special auxiliary CR the range of common auxiliary / CC common aux. connected with : CP common aux. connected with + M simple main number MS combined special auxiliary MSS combined with special auxiliary outside special auxiliary table MC combined with a common auxiliary MCS combined with common and special auxiliary MM two main numbers connected with /, : or + MMSC two or more main numbers combined with special and common auxilary and related to other numbers **NOTE:** a choice can be made to reduce the number or extend the number of types depending on the database tool		M	N
1.3	FACET TYPE	If UDC heading belongs to fundamental facets, this element will contain the code of facet. The function of this field is the revision of UDC towards fully faceted system. List of codes **A** Theory [this is not a fundamental facet but is needed for the UDC] **C** Time **D** Place **E** Agent **F** By-product **G** Product **H** Patient **J** Operation **K** Process **L** Material **M** Property **N** Part **P** Kind **S** Thing	CODED	O	R
1.4	TABLE	Codes the table of the UDC heading. List of codes: aTable Ia - Coordination. Extension	CODED	M	N

[189] This model makes a distinction between text and UDC notation (which also may contain symbols, numerals and letters). Elements that contain text have to be repeatable, are held in a separate table, and have a language type added to them.

	ELEMENT NAME	DEFINITION	VALUE TYPE[189]	MANDATORY/OPTIONAL	REPEATABLE/NON-REPEATABLE
		b Table Ib - Relation. Subgrouping. Order-fixing c Table Ic - Language d Table Id - Form e Table Ie - Place f Table If - Ethnic Grouping g Table Ig - Time h Table Ih - Notations from non-UDC sources i Table Ii - Viewpoint (cancelled 2000) k Table Ik - Persons and Materials l Section II. Special auxiliary subdivisions. M Main table			
1.5	SPECIAL AUXILIARY TYPE	If the UDC heading is a main number combined with a special auxiliary this codes the type of special auxiliary. List of codes A hyphen (-) auxiliary B point-nought (.0) auxiliary C apostrophe (') auxiliary D other, e.g. final digits (...1/...9) combine code with: RFlag! UDC heading contains application rule Note: If the UDC heading represents the beginning of a special auxiliary table that contains the rule for application of special auxiliary table - then add a flag 'R'. If this field contain R then the field of application note should be present.	CODED	O	R
1.6	APPLICABLE SPECIAL AUXILIARY TABLE	This element points to the beginning of the special auxiliary table(s) applicable for the UDC heading.	NOTATION	M	R
1.7	PARALLEL DERIVATION	If the UDC heading is derived by parallel division this field contain a UDC heading where instruction for derivation can be found.	NOTATION	O	N
1.8	BROADER CLASS	Contains the immediate broader class of the UDC heading	NOTATION	M	N
1.9	HIERARCHY CODE	This is an automatically generated code for the hierarchical level of the UDC heading and its sequence in sorting. The code is generated based on a broader class notation (e.g. '1.1.22' represents the 22nd UDC heading on the third hierarchical level of the first object). This code can be used for controlling the hierarchical level in interface display: expanding/collapsing hierarchies and for management of hierarchical levels for the purpose of export, distribution of producing a new edition.	TEXT	M	N
2	**DECOMPOSED UDC HEADING**	If the UDC heading is not a single main or a single common auxiliary, this field codes the components of the UDC notation. In the **UDC Research Database** decomposed elements are held in a different table which contains number/symbol type and building sequence code. For the purpose of information exchange a textual coding can be used to search pre-composed elements and for sorting of UDC numbers. UDC data can be exported containing this kind of coding Codes a Plus addition [+] b Stroke range [/] c Simple number d Intercalation	CODED NOTATION	M	R

	ELEMENT NAME	DEFINITION	VALUE TYPE[189]	MANDATORY/OPTIONAL	REPEATABLE/NON-REPEATABLE
		e Colon [:] combination f Double colon [::] combination g Combination Language [=...] h Combination with Form [(0...)] i Combination Place [(1/9)] j Combination with Ethnic Grouping [(=...)] k Combination with Time ["..."] l Combination with non-UDC notation [*] m Combination with A/Z n Combination with -05 Persons/Materials o Combination with -04 Processes p Combination with -03 Materials q Combination with -02 Properties r Combination with hyphen (-) special auxiliary s Combination with point-nought (.0) special auxiliary t Combination with apostrophe (') special auxiliary z precedes every classmark from the main table Note: Notation from main tables is preceded with z. Each component is separated by comma. UDC notation is entered without symbols.			
3	**PARALLEL DIVISION INSTRUCTIONS**	If parallel division can subdivide the UDC heading this field contains instructions for how this can be performed. If this field is present then the field of examples of parallel division has to be present. Field contains source notation and target notation	NOTATION	O	R
4	**CAPTION**	Textual description of a classmark. Further expanded through 'verbal examples'	**TEXT**	M	R
4.1	CAPTION VERBAL EXAMPLES	Includes examples of the core concepts. In compiling the MRF this field was used to include concepts from the lower level of hierarchy from the full edition. This field can still be used when producing abbreviated editions from the MRF	**TEXT**	O	R
5	**SCOPE NOTE**	Information about classification heading that explains the scope of the meaning i.e. its semantic coverage.	**TEXT**	O	R
5.1	SCOPE NOTE NOTATION VALIDATION	If scope note contains UDC notation this field will contain the UDC headings whose cancellation would affect the notation stated in the scope note.	NOTATION	O	R
6	**APPLICATION NOTE**	This note contains technical instructions regarding application or number building. This field often contains pre-synthesised notation and has to have a notation validation subfield	**TEXT**	O	R
6.1	APPLICATION NOTE NOTATION VALIDATION	If an application note contains UDC notation - this field will contain the UDC headings whose cancellation would affect the notation stated in the application note.	NOTATION	O	R
7	**EDITORIAL NOTE**	This note contains information about a UDC heading that publishers may want to display e.g. the source of vocabulary, the underlying policy in developing the class, pointers to other vocabularies, reference sources used, etc.	**TEXT**	O	R

	ELEMENT NAME	DEFINITION	VALUE TYPE[189]	MANDATORY/OPTIONAL	REPEATABLE/NON-REPEATABLE
8	**EXAMPLES OF COMBINATION**	Examples of combinations have to be held as decomposed elements and when exported they can be coded using the list of codes from the field of Decomposed_UDC_Heading. Examples of combination also have to have a field for notation validation with UDC headings, from which examples are derived.	NOTATION	O	R
8.1	EX. OF COMB. NOTATION VALIDATION	When an example of combination contains a combination of special auxiliary or notation resulting from parallel division - this field will contain UDC headings whose cancellation would affect the pre-combined notation used in the examples of combination.	NOTATION	O	R
8.2	EX. OF COMB. DESCRIPTION	An example of combination description follows each example of combination. Every notational example may have description in several languages	TEXT	O	R
9	**PARALLEL DIVISION EXAMPLE**	Contains an example of parallel division. This element should be populated when the element of Parallel Division Instruction is present.	NOTATION	O	R
9.1	PARALLEL DIVISION EXAMPLE DESCRIPTION	Description follows a parallel division example. Every notational example may have description in several languages	TEXT	O	R
9.2	PARALLEL DIVISION EXAMPLE VALIDATION	This field will contains UDC headings whose cancellation would affect the pre-combined notation used in the example of parallel division. This field is present whenever the field of Parallel Division Instruction is present.	NOTATION	O	R
10	**REFERENCE**	'See also' reference points to other related UDC headings in the system. In the process of display or export, the description of the reference is automatically generated from the caption of the referred heading. Reference notation can be validated against the UDC heading.	NOTATION	O	R
11	**KEYWORDS**	Lists by order of importance all the keywords relevant for the searching of UDC heading, irrespective whether these are present in the caption or not. Keyword may consist of more than one word. Field may be automatically generated from the caption and manually edited.	TEXT	M	R
12	**CHAIN INDEX**	Contains every keyword entry followed by the caption of the broader class. This can be generated automatically and gradually edited to create a full chain index	TEXT	O	R
13	**THESAURUS DESCRIPTOR**	This is a controlled term. It can be automatically generated from the first (most important) keyword but ideally it would be edited and created manually. In the process of creating a thesaurus this field has to be populated first	TEXT	O	N
13.1	THESAURUS USE FOR	Contains synonyms that may be created from keyword field or, ideally, manually assigned	TEXT	O	R
13.2	THESAURUS BROADER TERM	May contain descriptor denoted to the broader class or, ideally, manually assigned to a broader term	TEXT	O	N
13.3	THESAURUS NARROWER TERMS	May contain descriptors denoted in the narrower classes or, ideally, manually assigned terms	TEXT	O	R
13.4	THESAURUS RELATED TERMS	In general may contain thesaurus descriptors from any UDC heading mentioned in the field of reference or, ideally, manually assigned terms	TEXT	O	R

	ELEMENT NAME	DEFINITION	VALUE TYPE[189]	MANDATORY/OPTIONAL	REPEATABLE/NON-REPEATABLE
13.5	THESAURUS SCOPE NOTE	Explains the scope of descriptor	TEXT	O	N
14	**MAPPING TERM**	Contains term of an indexing language to which UDC heading can be mapped	NOTATION /OR TEXT	O	R
14.1	SYSTEM	Contain the name through which the system to which UDC is mapped can be identified	TEXT	O	R
14.2	MAPPING TYPE	This field could contain a controlled list of code that qualifiers the mapping. A possible list is: 1 from specific to (UDC heading = specific) 2 from broader to specific 3 partial semantic overlap 4 exact matching	CODED	O	R
15	**INTRODUCTION DATE**	This field contains the date of the introduction in the form YYMM	NUMERIC	M	N
15.1	INTRODUCTION SOURCE	Refers to instruction based on which the number was created. At present this is E&C publication	TEXT	M	N
15.2	INTRODUCTION COMMENT	If applicable this can contain further information on introduction	TEXT	O	N
15.3	REPLACING NUMBER	If the heading introduced is replacing other UDC heading(s) that were cancelled	NOTATION	O	R
16	**REVISION DATE**	Revision date. Database tool should hold the table with a strict chronological sequence so that the last revision and associated fields, source and comments can be traced	NUMERIC	O	R
16.1	REVISED FIELDS	Lists fields that were revised against the date	NUMERIC	O	R
16.2	REVISION COMMENTS	If applicable contains explanation on revision	TEXT	O	R
17	**CANCELLATION DATE**	This field contains data of cancellation in the form YYMM. If this field is present the record is logically deleted from the database	NUMERIC	O	N
17.2	CANCELLATION SOURCE	Refers to instruction based on which UDC heading is cancelled	TEXT	O	N
17.3	REPLACE BY	If applicable contains UDC heading that is replacing cancelled notation	NOTATION	O	R
18	EDITION TYPE	The database tool is used to hold more than one editions of the UDC e.g. UDC MRF and UDC extended edition or UDC pocket edition, this field could be used to identify edition. One UDC heading may belong to more than one edition.	TEXT	O	R
19	MAINTENANCE NOTES	This field is used by the reviser and database administrator to record work to be done. It can be used to temporary flag records listed for revision or temporary work.	TEXT	O	R
19.1	MAINTENANCE DATE	Date of the note	NUMERIC	O	R
19.2	SIGNATURE	Contains the name of the person making the note	TEXT	O	R

10.4 Range notation automation

classmark_id	classmark_tag
8858	(737.765/.783)

classmark_id	span_stem	span_type	span_prefix	first_span	last_span	span_suffix
8858	(737	N	.	765	783	)

classmark_id	seq_no	span_tag
8858	1	(737.765)
8858	2	(737.766)
8858	3	(737.767)
8858	4	(737.768)
8858	5	(737.769)
8858	6	(737.770)
8858	7	(737.771)
8858	8	(737.772)
8858	9	(737.773)
8858	10	(737.774)
8858	11	(737.775)
8858	12	(737.776)
8858	13	(737.777)
8858	14	(737.778)
8858	15	(737.779)
8858	16	(737.780)
8858	17	(737.781)
8858	18	(737.782)
8858	19	(737.783)

The above screen shots of SQL output from the UDC Research Database show three tables that detail the expansion of range notation into the schedules' full content. The first table shows the original notation. The second table shows how this notation is broken down to allow the expansion of the range and the third table shows the numbers generated that are contained in the UDC range notation.