TITLE-DERIVATIVE PRINTED INDEXES FOR ARABIC PERIODICAL ARTICLES

BY

IMAD A.M. BACHIR

A thesis submitted to the University of London in fulfilment of the requirement for the degree of Doctor of Philosophy in Library and Information Studies

School of Library, Archive and Information Studies
University of London
January 1995
Acknowledgments

I am grateful to Dr. Andrew Buxton for supervising this research, and for his continuous encouragement in carrying it out. I also thank the Lecturers and the secretaries at the School of Library, Archive and Information Studies for their kindness. The people with whom I have had useful communications and discussions, in particular my former teachers at the Lebanese University Mr 'Ahmad Talib, Dr Maude Estefan, Mrs 'Ayda Nu‘man and my professional colleague Mr 'Ubaydli 'Ubaydli are also thanked for their encouragement. My warm thanks are also extended to the staff and indexers of al-Fihrist Academic Research Institute for their assistance and for Mrs Helen Robbins, Head of Reader Services at BLISS, for proofreading the thesis.

My gratitude is also extended to the Hariri Foundation for supporting me during my study period in Britain and to the staff at London and Paris offices for their care.

Finally, I praise God for his grace and for giving me the strength and determination to complete this research.

I dedicate this work to my wife Batoul for her patience and encouragement and to my son Hassan for being born within the course of my study.

---------------------------------------------------------

Parts of this thesis have been published as:

Abstract

This thesis investigates the issue of using titles of Arabic academic periodical articles as a basis for printed indexes generated automatically by means of computers. The informativeness of the Arabic titles is examined by counting the number of their keywords in 16 disciplines and comparing the results with those for English periodical articles in the same subjects. Arabic titles are found to be as informative as English titles. The representation of the titles of their accompanying articles is examined by matching between the titles and their corresponding subject headings assigned by indexers in 10 subjects. To support the encouraging results of the title keywords and subject headings matching process, a separate study evaluates the representation of titles by comparing the keywords found in titles to the content of their accompanying topic sentences in the introduction and objectives parts of articles in 5 disciplines.

The practical side of the issue is also observed by the actual production of two types of printed indexes: one conventional, representing al-Fihrist which is the most used Arabic index in the Arab countries, and another unconventional KWOC type, representing the title keyword indexing technique with the assistance of the Arabic version of OMNIS 3.3. The retrieval effectiveness of the two types is examined by commissioning 20 Arab users to retrieve relevant articles to 9 queries composed to embody the various cases of retrieval difficulty.

The sample Arabic titles used in every examination process in the thesis are taken from current periodicals which use the Modern Standard Arabic as a medium of communication. Bearing in mind the retrieval problems and the
problems evolved during the production process of the Arabic KWOC type index, the thesis concludes that titles of Arabic articles in the examined disciplines can be used to produce acceptable printed indexes.
of periodicals..............44

2.2.5 Printed indexes to the contents of collections of periodicals
published by specific organizations.............47

2.2.6 Printed indexes to the contents of collections of periodicals
published in specific countries.............48

2.3 Arabic indexes to the contents of a group of periodicals.........49

2.3.1 The retrospective Arabic printed indexes.............50

2.3.1.A Alphabetical subject indexes.............51

2.3.1.B Classified indexes.............54

2.3.2 The current Arabic printed indexes.............57

2.3.2.A Ceased publication indexes.............57

2.3.2.A.1 The Analytical Index of Arabic Newspapers
and Journals.............58

2.3.2.A.2 The Educational Index of Periodical Articles in
the United Arab Republic.............59

2.3.2.B Current publication indexes.............63

2.3.2.B.1 The Analytical Index of Iraqi Journals.............64

2.3.2.B.2 Al-Fihrist.............65

2.3.2.B.3 The Analytical Index of the Arabic and Foreign Periodicals
and Bulletins Available at the Information and Documentation Centre
of the Gulf Cooperation Council Secretariat General .............69

2.3.2.B.4 The Islamic Index.............72

2.3.2.B.5 The Palestine Record.............76

2.4 Other printed bibliographical services.............79

2.5 Arabic subject heading lists and thesauri.............82

2.5.1 Arabic subject heading lists.............82

2.5.2 Arabic thesauri.............84
Chapter three: Keyword indexing techniques and the case for Arabic literature

3.1 Introduction

3.2 History of keyword indexing

3.2.1 Concordances

3.2.2 Abstracts

3.2.3 Titles

3.3 Computer characteristics

3.3.1 Speed

3.3.2 Accuracy

3.3.3 Reliability

3.3.4 Cost

3.4 Title-derivative indexes generated by computers

3.4.1 Indexes using the rotation of title KeyWords-In-and-Out-of-Context

3.4.1.1 KeyWord-In-Context (KWIC)

3.4.1.2 KeyWord-Out-of-Context (KWOC)

3.4.1.3 Word and Author Index (WADEX)

3.4.1.4 Author and KeyWord-In-Context (AKWIC)

3.4.1.5 Online KeyWord-In-Context (OKWIC)

3.4.2 Indexes using the rotation of enriched title KeyWords-In-and-Out-of-Context

3.4.2.1 KeyWord-With-Context (KWWC)

3.4.2.2 Enriched Title-based-KeyWord (ETKW)

3.4.3 Indexes employing title keywords coordination

3.4.3.1 Permuted Subject Index (PSI)
3.4.3.B DOuble-KWIC coordination index (DOKWIC) 115
3.4.3.C PANDEX index 115
3.5 Title keyword indexing and Arabic articles 119
3.6 Arabic title keyword indexing tools 119
3.7 Instructions to authors about titles 121
3.7.1 English language journals and books 121
3.7.2 Arabic language journals and books 125
3.8 Functions of titles of Arabic periodical articles 126
3.8.1 In the primary journals 127
3.8.1.A In content lists 127
3.8.1.B In reference lists 129
3.8.2 In the secondary services 129
3.8.2.A In the reproduction of contents pages 129
3.8.2.A.1 Contents of Arabic Periodicals (CAP) 130
3.8.2.A.2 In journals and daily newspapers 131
3.8.2.B In conventional indexes arranged by subject headings 131
3.8.2.C General and specialized bibliographies 131
3.8.2.D Titles indexes 132
3.8.2.E In machine-readable databases 132
3.9 Advantages and drawbacks of title keyword indexing 135
3.9.1 Advantages 135
3.9.2 Drawbacks 136
3.10 Conclusion 137

Part two: Suitability of titles of Arabic periodical articles for indexing
Chapter four: The information content of titles of Arabic articles and comparison with English titles 139
4.1 Literature survey..............140
4.1.1 English titles..............140
4.1.2 Arabic titles..............156
4.2 A study of the information content of Arabic titles..............157
4.2.1 Objectives of the study..............157
4.2.2 Method and samples..............157
4.2.3 Arabic stopwords list..............161
4.2.4 Results..............163
4.3 A study of the information content of English titles..............165
4.3.1 Objectives of the study..............165
4.3.2 Method and samples..............165
4.3.3 Results..............166
4.4 Comparison between the results of the two studies using the T test..............168
4.5 Discussion of the results for the two studies..............170
4.6 Conclusion..............172

Chapter five: Matching title keywords of Arabic articles and their corresponding subject headings in an interdisciplinary printed index..............173
5.1 Introduction..............174
5.2 Previous studies..............174
5.3 Interdisciplinary studies..............176
5.4 Criticism of these studies..............178
5.5 Criticism of the criticism..............180
5.6 New study for titles of Arabic periodical articles..............181
5.7 Objectives of the study..............181
5.8 Method and samples..............182
5.8.1 Sample determination ..........182
5.8.2 Categories of analysis ..........184
5.8.3 Definitions of categories with Arabic examples ..........185
5.8.3.A Exact match : entire heading ..........185
5.8.3.B Exact match : access point only ..........185
5.8.3.C Stem match : access point ..........186
5.8.3.D Synonym match : access point ..........186
5.8.3.E No match ..........187
5.8.3.F Subdivisions match ..........187
5.9 Results and analysis ..........189
5.10 Discussion ..........196
5.10.1 Single categories ..........196
5.10.1.A Exact match : entire heading ..........196
5.10.1.B Exact match : access point only ..........196
5.10.1.C Stem match : access point ..........198
5.10.1.D Synonym Match : access point ..........198
5.10.1.E No match ..........198
5.10.1.F Subdivisions match ..........199
5.10.2 Combined categories ..........199
5.11 Conclusion ..........203

Chapter Six : The occurrence of title keywords in their article's topic sentences ..........206
6.1 Introduction ..........207
6.2 Objectives ..........207
6.3 Usage of topic sentences ..........208
6.3.1 The use of topic sentences in this study ..........208
6.3.2 Previous usage of topic sentences ..........210
6.4 Method and samples............213
6.4.1 The location of topic sentences............213
6.4.2 The matching criteria............214
6.4.3 The recognition of anaphors............215
6.4.4 Samples and data gathering............218
6.5 Results............219
6.6 Discussion............227
6.7 Conclusion............233

Part three: Arabic language characteristics and the efficiency of title-derivative printed indexes

Chapter Seven: Producing an Arabic title-derivative printed index and users' evaluation............235
7.1 Bibliographic information systems and Arabization............236
7.2 The use of computers to produce Arabic printed indexes............239
7.3 Title keword derivation techniques and the Arabic language............241
7.3.1 The rotation technique and the Arabic language............242
7.3.2 The permutation technique and the Arabic language............243
7.4 The title-derivative index types and the Arabic language............243
7.4.1 The KWIC type and the Arabic language............244
7.4.2 The KWOC type and the Arabic language............248
7.4.3 The PSI type and the Arabic language............249
7.5 Producing a title-derivative printed index for Arabic articles............251
7.5.1 Choosing the software............251
7.5.2 The production process of the KWOC type index............257
7.5.3 Problems involved in the production process............261
7.5.3.A The inseparable prepositions............261
Chapter seven: Characteristics of the Arabic language and problems of retrieval from an Arabic title keyword printed index

7.5.3.B The various types of the Arabic ‘alif character

7.5.3.C Voweling marks

7.5.3.D Indian numbers

7.6 Comparison of two types of Arabic printed indexes

7.6.1 The evaluation process

7.6.2 Producing an al-Fihrist type comparable index

7.7 Methodology of the comparison between al-Fihrist and the KWOC types

7.8 Results of the comparison

7.8.1 The al-Fihrist type index (category A)

7.8.2 The Arabic KWOC type index (category B)

7.9 Discussion

7.10 Conclusion of the comparison process

Chapter eight: Characteristics of the Arabic language and problems of retrieval from an Arabic title keyword printed index

8.1 Introduction

8.2 Arabic alphabet

8.3 Arabic language features

8.4 Retrieval problems

8.4.1 Retrieval problems handled at the input stage

8.4.1.A Usage of voweling marks

8.4.1.B Inseparable prepositions

8.4.1.C Variation in the orthography of some letters

8.4.2 Retrieval problems requiring sophisticated software

8.4.2.A The definite article

8.4.2.B Existence of synonymous keywords

8.4.2.C Existence of plural and singular forms
Chapter nine : Conclusions and recommendations

9.1 Background to the study

9.2 Procedures followed in the thesis

9.3 Findings of the thesis

9.3.1 Arabic periodical indexes

9.3.2 Titles of Arabic articles

9.3.3 Informativeness of Arabic titles

9.3.4 Representativeness of Arabic titles

9.3.5 Arabic KWOC evaluation

9.4 Conclusions

9.5 Recommendations

9.6 Future studies

Appendices and references

Appendix 1.1 : The Library of Congress transliteration system for Arabic letters

Appendix 1.2 : List of the Arabic indexing tools which were mentioned in the thesis and their English transliteration

Appendix 4.1 : List of the Arabic periodicals which were used to study the informativeness of Arabic article titles

Appendix 4.2 : The stopwords list which was used to study the information content of titles of Arabic periodical articles

Appendix 4.3 : List of the English periodicals which were used to study the informativeness of English article titles
Appendix 5.1: List of the Arabic periodicals used in the matching process between titles of Arabic articles and their corresponding subject headings.............335

Appendix 6.1: List of the Arabic phrases found at the beginning of the examined Arabic topic sentences.............336

Appendix 6.2: List of the Arabic periodicals from which the samples of titles and topic sentences were selected.............337

Appendix 7.1: Copy of the form which was distributed to the searchers of al-Fihrist type index.............338

Appendix 7.2: Copy of the form which was distributed to the searchers of the Arabic KWOC type index.............343

Appendix 8.1: The Arabic characters and their various shapes within a word, taken from Wright (1933).............348

Reference list.............349
Tables

Table 2.1: number of periodicals in each Arab Gulf State alongside the number of individual printed indexes which they publish..........43

Table 4.1: composite of the prior researches which examined the information content of titles of English periodical articles..........141

Table 4.2: the number of substantive words found in titles of chemical articles during the selected years - Tocatlian (1970)..............143

Table 4.3: the difference in means found between the selected years - Tocatlian (1970)...............143

Table 4.4: average of titles length for the four selected journals - Bird and Knight (1975)..............145

Table 4.5: average of titles length in Nature - Bird and Knight (1975).............145

Table 4.6: average of titles length in chemical literature - Bird and Knight (1975).............145

Table 4.7: Buxton and Meadows's (1977) results for titles in scientific journals.............148

Table 4.8: Buxton and Meadows's (1977) results for titles in non-scientific journals.............148

Table 4.9: Balog's (1981) results for titles in New Zealand Journal of Agricultural Research.............150

Table 4.10: Diodato's (1982) results for titles in five selected journals.....151

Table 4.11: sample of an Arabic keyword with three inseparable prepositions.............161

Table 4.12: the mean and standard deviation values of contents of Arabic scientific article titles in five disciplines.............163

Table 4.13: the mean and standard deviation values of contents of Arabic non-scientific article titles in eleven disciplines.............164

Table 4.14: the mean and standard deviation values of contents of English scientific article titles in five disciplines.............166
Table 4.15: the mean and standard deviation values of contents of English non-scientific article titles in eleven disciplines.............167

Table 4.16: values of T for differences between scientific titles of English and Arabic articles.............168

Table 4.17: values of T for differences between non-scientific titles of English and Arabic articles.............169

Table 5.1: Sedano's (1964) results of comparisons of title words and subject headings.............177

Table 5.2: Lane's (1964) results of comparisons of title words and subject headings.............178

Table 5.3: number of sample articles and their allocated subject headings in 10 subjects taken from al-Fihrist.............183

Table 5.4: number and percentage of access points and subdivisions for subject headings which matched keywords in titles of Arabic articles.............191

Table 5.5: overall number of access points which matched keywords in Arabic titles and their percentage.............193

Table 5.6: number and percentage of similar words between title keywords of Arabic articles and their corresponding subject headings in 10 subjects.............194

Table 5.7: average number of keywords per Arabic title and subject headings per article in 10 fields.............195

Table 5.8: percentage of similarities between title keywords and words in subject headings in 10 fields.............201

Table 5.9: number and percentage of subject headings according to categories of matches.............202

Table 5.10: number and percentage of matches according to their discipline groups.............203

Table 6.1: number of sample, substantive words in topic sentences; substantive words in titles; substantive words which matched; proportion of substantive words which matched; and proportion of substantive words in titles of agriculture.............220
Table 6.2: number of sample, substantive words in topic sentences; substantive words in titles; substantive words which matched; proportion of substantive words which matched; and proportion of substantive words in titles of library and information science........221

Table 6.3: number of sample, substantive words in topic sentences; substantive words in titles; substantive words which matched; proportion of substantive words which matched; and proportion of substantive words in titles of linguistics........222

Table 6.4: number of sample, substantive words in topic sentences; substantive words in titles; substantive words which matched; proportion of substantive words which matched; and proportion of substantive words in titles of philosophy........223

Table 6.5: number of sample, substantive words in topic sentences; substantive words in titles; substantive words which matched; proportion of substantive words which matched; and proportion of substantive words in titles of sociology........224

Table 6.6: mean and standard deviation value of the relative matching in titles of each of the examined subjects, arranged according to their mean values........226

Table 6.7: mean and standard deviation values of the absolute matching in titles; substantive words in titles; and substantive words in topic sentences, arranged according to the mean values of their absolute matching........226

Table 6.8: number and percentage of representative and non-representative titles in the five selected subjects........230

Table 7.1: the average recall and precision of the al-Fihrist type searchers category........278

Table 7.2: the average time taken by searchers of the al-Fihrist type category........279

Table 7.3: the average recall and precision of the KWOC type searchers category........281

Table 7.4: average time taken by searchers of the KWOC type category...282

Table 7.5: the average recall and precision achieved by searchers of the al-Fihrist type and the KWOC type at the query level..........288
Figures

Figure 2.1: sample of the Index of Articles in Saudi Periodicals specialized in Libraries and Information.................52
Figure 2.2: sample of Index Arabicus..................56
Figure 2.3: sample of the Analytical Index of Arabic Newspapers and Journals..............61
Figure 2.4: sample of the Educational Index of Periodical Articles in the United Arabic Republic..............62
Figure 2.5: sample of the al-Fihrist..................71
Figure 2.6: sample of the Islamic Index.................75
Figure 2.7: sample of the Palestine Record.............78
Figure 2.8: sample of the bibliography provided by the Arab Future Journal.............81
Figure 3.1: sample of Watt's Bibliothica Britannica, taken from Craven (1986)..............95
Figure 3.2: sample of the Manchester Public Library catalogue by Crestadoro (1856)..............96
Figure 3.3: the generation process of a KWIC index............103
Figure 3.4: sample of a KWIC type index..............104
Figure 3.5: sample of a KWOC type index with indexing window at the left margin.............105
Figure 3.6: sample of a KWOC type index with indexing window above the index line entry.........106
Figure 3.7: sample of a WADEX type index.............107
Figure 3.8: sample of an AKWIC type index.............108
Figure 3.9: sample of an OKWIC type index.............109
Figure 3.10: sample of a KWWC type index.............112
Figure 3.11: sample of an ETKW type index showing the process of title enrichment.............113
Figure 3.12: sample of the final output of an ETKW type index, taken from Rajendran (1986).............113
Figure 3.13: sample of a PSI type index..............116
Figure 3.14: sample of a DOKWIC type index...........117
Figure 3.15: sample of a PANDEX type index...........118
Figure 3.16: sample of "IBN ALNADEEM" search result........120
Figure 3.17: samples of instructions to authors on the composition of titles, collected by Feinberg (1973).............124
Figure 3.18: cover of the Arabic Journal of Information Science showing titles of articles in the issue........128
Figure 3.19: sample of the Bibliography of the Islamic Maghrebin.............133
Figure 3.20: sample of the Message of the Library's title index.............134
Figure 7.1: sample of Index of Events Section in the Panorama of Events Journal.............245
Figure 7.2: Kasem's manual KWIC index.............246
Figure 7.3: Salem's KWIC list of Arabic medical terminology ..........247
Figure 7.4: sample of the KWOC index to the al-Jami’a thesaurus ..........250
Figure 7.5: sample of KWIC type index produced by Dobis system ..........253
Figure 7.6: sample of the Arabic KWOC type index produced for this study ..........260
Figure 7.7: sample of the al-Fihrist type index produced for this study ..........270
Figure 8.1: sample of al-Fihrist index showing the use of scope notes beside ambiguous subject headings ..........299
Figure 8.2: sample of the Islamic Index showing the use of scope notes beside ambiguous keywords ..........300
Figure 8.3: Arabic MS-DOS code page 708 (compatible with ASMO 449 and 708) showing the various types of ’alif ..........305
Chapter one

Introduction

This thesis is devoted to investigating a vital issue related to the indexing of the contents of Arabic articles published in academic and research periodicals. This issue comprises the possibility of using one of the unconventional indexing techniques known as title keyword indexing to overcome the deficiency of the current indexing tools in analysing the contents of Arabic periodical articles. However, before getting to the purpose of this research and prior to the clarification of some related aspects to its topic, it is expedient to define what is a printed index, and what is meant by conventional and unconventional periodical indexing.
1.1 What is a periodical printed index?

Wagner (1960), Harrod (1977, p.412), the American Library Association (1983, p.116), British Standards Institution (1988, p.2), and Cleveland (1990, p.26) generally agree that an index is a systematic guide to the contents of items or documents: it guides users to the items that they are looking for or interested in by means of keywords or subject headings arranged in a known order. This systematic guide offers access to the huge volume of literature by communicating the existence of items in the current literature from one side and by making past literature available through facilitation of retrospective searching from the other side. Thus, a periodical index can be defined as an orderly bibliographic tool to the intellectual content and physical location of indexed articles. It is only a guide and as such it does not provide the desired article itself, but it employs a set of bibliographic tags which refer to the source of information for which the user is searching. Printed periodical indexes are of many kinds. However, this thesis is looking into one type of them often called title-derivative printed indexes, which normally employ the rotation and the permutation techniques in the process of their generation.

As for the basic components of such printed indexes, Azgaldov (1969) and Campey (1972, p.3) gave a considerable description of printed subject indexes, but it is convenient in this section to mention only those of derived printed indexes, which are the following:

a - The entry point part (title keywords).

b - The context in which keywords are being considered (the titles)

c - Author and bibliographic details which are related to the indexed article.
1.2 Conventional and unconventional periodical indexing

In contrast with conventional indexing, in which indexing terms are not necessarily selected from the text of an article but often from either a freely chosen vocabulary or from standardized lists, title-derivative indexing involves the extraction of indexing terms from the article titles. In English literature, specialists often refer to conventional printed indexes as conventional or assigned printed indexes and to title-derivative indexes as unconventional or derived printed indexes. A main difference between these two types of indexing is that the conventional indexing requires the use of indexer efforts in conducting the intellectual and clerical tasks needed to produce a printed index. That is to say, a trained indexer is always expected to read or scan the article in order to assess its subject matter, then to assign terms or concepts that will serve to identify the article for later retrieval. The indexer efforts may also be needed to carry out the clerical task, if a computer is not used for this operation. The clerical task involves the arrangement and the display of the index entries according to a specified format. Whereas, the title-derivative indexing process depends to a large extent on the computer to carry out both the intellectual and clerical operations.
1.3 Incentives of this research

Most of the Arabic periodical indexing systems which are being used to analyse the growing numbers of Arabic periodical articles are conventional. Only few have recently started to rely on computer facilities to produce printed indexes with the subject headings assigned by an indexer, e.g. the Islamic Index (الكشف الإسلامي). In such cases computer facilities are used only to perform purely clerical operations like storing the data and then sorting them out according to a specific order. As a matter of fact, it has not been until recently that indexing services have been available for the contents of Arabic periodicals published in the Arabic language. In the early 1980s conventional indexing services for current Arabic periodicals were introduced with the publication of individual periodical indexes by a number of research periodicals such as World of Thought (عالم الفكر); Arab Gulf Journal (مجلة الخليج العربي); and the Jordanian periodical Studies (دراسات).

A proper index for Pan-Arabic periodicals was not begun until 1981 with the publication of al-Fihrist (الفهرست). It is a quarterly index to the contents of academic and research Arabic periodicals ranging from monthlies to annuals. Despite the fact that many trials to produce indexes for Arabic periodicals were carried out before its appearance, al-Fihrist is widely considered as the first effort of its kind in the Arab world. Although the publication of al-Fihrist does fill a serious gap in existing Arabic bibliography, it is not a perfect tool to index the contents of Arabic periodicals and has not succeeded in being current. Some of the reasons behind this failure are the shortage of professional Arab indexers and the fact that conventional indexing consumes time and is expensive to keep updated.

The non-currency of the available indexing services for Arabic
periodicals on one side and the inadequate coverage of the Middle East periodicals by major indexing services published by Western institutions on the other have enhanced the feelings of the great need for a better indexing system and new methods to handle current Arabic literature. The inadequate and unsatisfactory coverage of Western indexing services to the literature published in Islamic countries in general and in the Arab countries in particular was revealed by Sattar and Rehman (1985) who assessed the degree of bibliographical control of Islamic literature in the coverage provided by four Western indexing services. These are SSCI (Social Sciences Citation Index) which is produced by the Institute of Scientific Information and has international coverage in the area of social sciences; SI (Social Sciences Index) and HI (Humanities Index) which are products of the H. W. Wilson Company, and complement each other in coverage; and finally the II (Index Islamicus) which is published by Mansell and compiled under the auspices of the Royal Academy for Islamic Civilization Research (Al al-Bait Foundation) and claims to cover the Islamic literature exclusively. In this domain, it can be mentioned that Arabic language is among Wellisch's (1980) list of non-Roman script languages whose literature has poor bibliographic control. Also, Yaghmai et al (1986) who assessed the bibliographical coverage of materials dealing with the Arabic language and Islamic cultures in 16 machine readable databases, stated that "any study of the Arab language and Islamic cultures will most assuredly be hampered by an inadequate access to information". Prior to that, Anwar (1983) mentioned that as far as the bibliographic control of information produced about Muslim countries (including the Arab world) is concerned, it is perhaps the weakest for any region.
1.4 Purpose of this research

As a remedy for the poor bibliographic control with regard to Arabic periodicals in the Arab world and to the weakness and inadequacy of the current Arabic periodical indexing tools, this thesis is investigating the possibility and relevance of using unconventional indexing techniques to enable users to satisfactorily keep abreast and to assist publishers in overcoming the existing limitations of current Arabic periodical indexing. One sort of unconventional indexing techniques which can be employed in such circumstances is the use of keywords of article titles to produce title-derivative printed indexes.

It is generally known that one part of the article in which the author usually attempts to indicate the subject matter is the title. However, this function of the title is sometimes used improperly by authors who are exploiting it in order to tempt readers by providing attractive but unindicative titles. The objective of this research is to provide an answer to the question of whether titles of Arabic periodical articles can be relied on as a basis for title word indexing techniques. To investigate this possibility, the thesis assumes that restrictions which formerly prevented the use of computers to generate automatic title-derivative printed indexes are diminishing, and claims that if titles of Arabic periodical articles are found adequate, then the use of such indexing techniques will enhance the dissemination of Arabic periodical literature to a larger audience.

1.5 Article titles and their importance

Having declared that the purpose of this thesis is to examine the validity of depending on titles of Arabic periodical articles to produce printed indexes
by means of computer, it is relevant to cite some of the definitions and quotations concerning the titles of academic articles and show their vital importance in information retrieval processes.

Rowley (1992, p.145) defines the title of a work as a "word, phrase, character, or group of characters, normally appearing in an item, naming the item or the work contained in it". Thus the title is a label associated with a work, especially if this work is a journal article. It is the first thing that we read when we scan a periodical's list of contents, and as Bottle and Seeley (1970) say "the title of a document or paper is probably the first thing about that document or paper that really registers in our minds unless one of the authors has an exceptionally well-known name".

The title of a journal article is the most important among other bibliographic information contained in articles as printed in the primary journals. These other elements could be:

a - Author's name and affiliation
b - Abstracts
c - Keywords
d - References

It is more important because it is usually more indicative of the content of its corresponding article than the author's name, which is not known in most cases. Unlike the abstract, keywords, and references, the title of an article always accompanies the article, and is one of the items of bibliographic information in both primary and secondary printed journals.

Apart from letters and short notice documents, it is hard to come across an article in a journal without a title, whereas articles that have no associated abstracts, no keywords or even no references are often found in some of the current journals and conference proceedings. Huth (1982, p.76) confirms this
in his chapter "writing the first draft". He says "all journal papers carry titles, and most include abstracts". Moreover, an author of a paper might be lucky to find a journal that considers his abstractless or referenceless paper for publication. But his luck will not help him in getting a journal which accepts an article without a title for publication. Mullins (1977, p.248) asks authors to make the titles of their papers clear and descriptive so that the editor will be helped to choose the appropriate referees.

In fact, the importance of the title is acknowledged by the authors themselves. It is in this part, which is a key element for the readers, that they usually attempt to indicate the subject matter of the content of their articles. However, it could be used improperly by some authors who exploit it in order to tempt the readers by providing attractive but unindicative titles. It is a key element to the readers because it is on the basis of the title that they frequently make decisions on whether or not to read the article's abstract, look at the list of references, or skim through its full text. It is believed that the cases where readers make further reference to the journal article on the basis of its author's name are fewer than those in which readers usually would find out "who has written a specific article" rather than "what has been written by a specific author". There are in fact indications that in many situations a reader is likely to remember a title better (Ayres et al, 1968).

In addition to this, the title of an article plays a very crucial role in the process of information transfer regardless of whether this occurs in its positive or negative sense. Bottle and Seeley (1970) stated that "information transfer in a positive sense will occur when the title contains words or word combinations that alert us to the possibility that the document may contain information of interest to us" and "our rejection of titles on the ground of lack of relevance of interest constitutes a filter through which passes the flood
of documents that we do not require. Thus for the majority of documents, the
titles of which we scan, information transfer in a negative sense occurs
through their titles ". Many scholars will not read articles whose titles do not
show clearly that the subject matter is related to their interests.

1.6 Rotation and permutation techniques

It was mentioned earlier in this chapter that title-derivative printed
indexes normally employ rotation and permutation techniques in the process
of their generation by means of computer. Since a confusion is found in
English and Arabic indexing literature regarding these two techniques, it is
useful to clarify this issue before getting to the various types of
title-derivative printed indexes which use them.

In rotated printed indexes there are as many access points as there are
keywords in the title. By keyword is meant every significant or substantive
word found in titles of the indexed articles. Section 3.1 of chapter 3 reveals
the various terms used in English to express a keyword. The combination
order of the title's keywords is retained, but index entries are made under
each keyword. The following example of a four keywords in an article title
represented in the four Arabic alphabet characters : 'alif (ی), b (ب), t (ت) and
th (ث) shows that the filing keyword is in each case the one in the same
indexing window or position as the Arabic letter 'alif (ی) in the first, but the
relative position of each keyword with respect to the others remains
unchanged.
Whereas in permuted printed indexes, all possible permutations of the title's keywords are used as access points. To give an example in symbolic form, if the same keywords as above are chosen 'alif (أ)', b (ب), t (ت) and th (ث) then reference to them would be from the following index entries:

The number of permutations that can be made of n keywords is n! (factorial n). For the above symbolic title keywords which are 4, n! is 24 permutations.

It is important not to confuse these two techniques with each other, an error into which many Arab and English authors and publishers have fallen. For instance several authors referred to KWIC (KeyWord-In-Context) and KWOC (KeyWord-Out-of-Context) indexing type as permuted indexes, when
in fact they are rotated indexes. For more details about such confusion, reference to Kemp et al (1972) and Foskett (1982, p.97) is advised.

1.7 Components of this thesis

This thesis is structured to consider the deficiency of Arabic periodical indexing and the solution to this problem. It consists of three parts.

The first part contains three chapters and provides a thorough and comprehensive introduction to the issue of Arabic periodical indexing and the hurdles to the objectives of the printed Arabic periodical indexes to be current and up-to-date indexing tools. It also intends to give a general overview of title keyword indexing and the ways in which the various types of title-derivative printed indexes are produced.

The second chapter of part one represents a state of the art report for Arabic indexing and Arabic printed indexes. It casts light on the history of Arabic periodical indexing and surveys the indexing products including the printed indexes which have ceased publication as well as the current ones. It concentrates on the current Arabic periodical indexes to show their deficiency and inadequacy in covering the contents of Arabic periodical articles and reveals the reasons behind this deficiency to conclude that a new indexing system is now needed.

The third chapter of part one defines the proposed title-derivative indexing techniques and reviews their history before and after the emergence of the computer as an indexing tool. It also reviews the various types of derived printed indexes and the way they are produced. In addition to that, it reveals the functions of article titles and their various use in the different types of Arabic bibliographical tools. It discusses the available literature for both Western and Arab authors with regard to title construction for their
articles, and ends with a condensed review of what has been said about the advantages and drawbacks of title keyword indexing techniques.

Part two of this thesis also consists of three chapters. It aims to provide the answer to the issue of the informativeness and representativeness of Arabic article titles. Since the proposed indexing method depends on keywords found in titles of Arabic articles, the quality of such printed indexes relies to a great extent on how well the titles of Arabic periodical articles are constructed by their authors. Three consecutive chapters are devoted to investigating the informativeness and representativeness of Arabic article titles. Informativeness is different from representativeness and informative titles may not be representative, a difference that was not noticed in many of the previous studies which dealt with the relevance of article titles for indexing.

Chapter four of this thesis is devoted to the study of the information content of titles of Arabic periodical articles and aims to find out how well Arab authors compose the titles for their articles in terms of the number of informative words. It determines the average length of Arabic article titles and the number of substantive words they include in contrast with the number of non-substantive words in 16 various fields. A similar study of the information content of English article titles was carried out for the purpose of comparison. English periodical articles have already been investigated and claimed to be informative according to previous studies.

Chapter five of this thesis is the first of two chapters devoted to the study of the titles' representation of their accompanying articles. To investigate the similarities between keywords used by authors in their article titles and by indexers in their subject headings, it examines the matching between the content of titles of Arabic articles and the contents of the
indexer-assigned subject headings for the same articles in 10 various fields taken from al-Fihrist (الفهرست) index. Another objective for this chapter is to determine the average number of subject headings assigned to Arabic articles and compare it with the average number of access points (keywords) found in the titles of the same Arabic articles.

Chapter six is also devoted to the examination of titles' representation of their accompanying articles. To overcome the criticism of the previous titles' representation study in chapter five, this chapter tries to follow a completely different approach. For this, it assumes that authors normally express the subject of their articles in what is known as the "topic sentence", often found in the "introduction" or "objectives" parts of an article. This chapter matches the keywords found in titles and the keywords found in the topic sentences manually extracted from the selected articles. It starts with defining the expression "topic sentence" and the boundaries of its use as well as reviewing its previous usage in the library and information science literature. Then the matching process is carried out with attention to other aspects that may affect the results and the chapter ends with a conclusion to the whole process.

The last part of the thesis also consists of three chapters. This part represents the practical side of the research, and like previous parts it includes studies and reveals facts regarding the Arabic title-derivative indexing, which have not been mentioned before. Chapter seven is devoted to revealing the problems involved in the process of using keywords of Arabic titles to generate automatic printed indexes, and to studying the user reactions to such types of Arabic printed indexes in contrast with that of conventional indexes. For this, the chapter describes the process of producing two different types of Arabic printed indexes. The first is a title-derivative printed index and the
latter is a conventional index to the same articles. These two types are then shown to a number of Arab users in a process to study the retrieval efficiency of the first in comparison to the conventional index. In this chapter, consideration is given to the Arabization activities for computer programs and the use of Arabized computer packages in librarianship and information work in the Arab institutions inside and outside the Arab countries.

Chapter eight is allocated to the study of the characteristics of the Arabic language and the problems involved if users depend on Arabic title-derivative printed indexes to search for their required articles. It provides some suggestions in this domain.

The final chapter is devoted to the conclusion of the research and includes some recommendations to authors and publishers regarding the construction of titles of Arabic periodical articles.
Chapter two

Arabic literature indexing and
Arabic periodical printed indexes

This chapter considers the state of the art of the current Arabic periodical indexing as well as printed indexes for the contents of Arabic periodical articles. It reviews the history of indexing activities in the Arab world and discusses the validity and sufficiency of current Arabic indexing techniques in covering and analysing the contents of academic and research periodical articles.
2.1 Introduction

Before starting the next part which is concerned with Arabic periodical printed indexes, it is relevant to mention that cooperation among Arab countries to produce general and specialized regional bibliographies was not remarkable until late 1972, when ALECSO (the Arab League for Education, Culture and Sciences Organization) published the first issue of the Arabic Publications Bulletin (النشرة العربية للمطبوعات). This bulletin constitutes an accumulation and coordination for the contents of other available Arabic national bibliographies as well as bibliographic descriptions and details on publishing activities given by Arab countries which do not have national bibliographies. This unprecedented effort in the Arab world is not ideal and still carries some drawbacks in its components. It only covers books and periodical articles and neglects other types of materials. Also it does not report the publishing activities in all Arab countries. In addition, the irregularity of its publication has turned it into more of a retrospective than a current bibliographic tool. The progress of this regional bulletin depends to a great extent on the commitment to the legal deposit system in the countries where it exists, or perhaps on making this system compulsory to authors, as well as the establishing of national libraries in the countries which are dropping behind. One more condition towards the development of such tools is the standardization of rules for the preparation of national bibliographies in the Arab States.

The remaining part of this chapter is not going to explore more such recent bibliographical works, and will skip the Arabic national and other specialized bibliographies because their coverage extends to monographs and other materials in certain cases and the fact that these bibliographies are
mostly annotated. In addition to that, most of these bibliographies are retrospective and do not have the characteristics of periodical article indexes which form the core of this thesis.

2.2 History of Arabic periodical indexing

The process of indexing the content of Arabic periodical articles is relatively new to the Arab world and the analysis and recording of most indexing products in the Arab world have been almost entirely neglected. The number of studies and essays by Arab authors which evaluate such products as to their efficiency and performance is almost nothing in comparison to other studies done on Western products by Western specialists. The bibliographical works of compilers interested in Arabic librarianship - Abdul Huq and Aman (1977) who cited 1475 titles, Pantelidis (1979) who cited 1047 titles, ‘Abd al-Hadi (1981, 1987, 1988, 1989a, 1989b,) who cited almost 10377 titles, Anwar (1985) who cited 338 titles in Arabic librarianship and information studies in the Arab world since 1900 until 1983, Tameem (1988) who cited 71 titles, and Mustafa (1991) who cited 349 titles in universities and institutional libraries - were consulted for studies on Arabic periodical indexing for that period. These bibliographies cover all possible printed materials including books, periodical articles, dissertations, theses, pamphlets and unpublished special mission reports.

As to the remaining period, other bibliographical services covering the Arabic literature in library and information studies were consulted from time to time. These services are offered by few current Arabic periodicals such as World of Books (عالم الكتب) and World of the Book. Specialized databases such as the LIBINFSC file on Dialog were also consulted for studies that dealt with the same topic.
The result showed that Arab and non-Arab authors had little interest in
this topic at the beginning and that their interest, although unremarkable,
started to be seen through the bibliographic records of the 1980s onwards.

Hamdy (1980) refers the desire for progress on the front of Arabic
bibliography and its controlling tools for many reasons. Some of these were:
the involvement of UNESCO (the United Nations Educational, Scientific,
and Cultural Organization) and other national and international organizations
in developing the library systems in the Arab world and then the return of
several Arab librarians trained in Europe and the United States to their native
countries. This need for controlling tools was expressed in local meetings,
national seminars, and regional conferences on libraries and librarianship in
the Arab world. Three of these early seminars in particular, one held in Beirut
1959, one held in Cairo 1962, and one held in Riyadh 1973, discussed the
problems of bibliographic control of Arabic materials and the need for
professional tools, which is one of the major problems standing in the way of
library and information services development in the Arab world. In addition
to other major librarianship issues, all three seminars recommended
establishing tools and standards for the production of indexes.

In the Arabic literature of library and information science, the Arabic
word kashshaf (كشاف) is used to match the English term (index) and the
word fahrass (فهرس) is used to express the meaning of the English word
(catalogue). Musa (1973, p.2) and al-Hajrasy (1974a, p.82) mentioned that in
the literature of librarianship in the Arab world the convention has settled on
the usage of the Arabic word kashshaf (كشاف) to be equivalent to the English
term (index).

The confusion in using both Arabic terms index (كشاف) and catalogue
(فهرس) to describe bibliographical works is reflected in the Arabic literature.
Some specialists use the word fahrass (catalogue, فهرس) to describe the product of indexing the contents of one or many journals. Some others use the word kashshaf (index, كشاف) to describe similar products. The word fahrass (catalogue, فهرس) has been in use for long in Arabic literature concerned with bibliography. It is an Arabization of the Persian term fihrist (فهرست) which was used in the title of the oldest Arabic bibliography "فهرست التراث" compiled by Ibn al-Nadim. Whereas the word kashshaf (index, كشاف) is recent in Arabic and according to the published Arabic indexes, the first bibliographic product to use this word is "the Educational Index of Periodical Articles in the United Arab Republic" (الكشاف التربوي لقاءات الدوريات في الجمهورية العربية المتحدة) launched in 1961 by the Centre for Educational Documentation as a semi-annual index and ceased in 1967. For the history of indexing and the printed indexes in the Western world it is advised to see Cornog (1983) and Wellisch (1983) who wrote extensive articles about this matter.

The confusion between what is an index and what is a catalogue is not Arab made. In the literature of library and information studies, a distinction is sometimes made between the two terms subject indexing and subject cataloguing. Thus, the same distinction is made between the output of the first process and the output of the latter. Usually the first output is called an index whereas the latter is called a catalogue. Lancaster (1991, p.14) stated that "subject cataloguing usually refers to the assignment of subject headings to represent the overall contents of complete bibliographic items within the catalogue of a library. Subject indexing is a term used more loosely; it may refer to the representation of the subject matter of parts of complete bibliographic items as in the case of an index at the back of a book". He sees this terminological distinction as meaningless, artificial, misleading,
inconsistent and only serves to cause confusion. He furthermore believes that "the process by which the subject matter of bibliographic items is represented in published databases printed or machine-readable form is almost invariably referred to as subject indexing, whether overall items or their parts are being discussed". ‘Abd al-Hadi (1982, p.18), an Arab scholar in this field described the difficulty of distinguishing between these two processes, although there is a trend to use the term subject cataloguing (فهرسة) to represent the subject matter of complete bibliographic items and to use the term subject indexing (تكتشيف) to describe the contents of parts which form larger or complete items. ‘Abd al-Hadi added that if the catalogue guides the user to both the bibliographic descriptions and subject cataloguing as well as to the location of the book in a library collection, the index reveals the analysed contents of a single or many books in specific groups or to one or more periodicals. On this interpretation, he believes that the difference between an index and a catalogue is a difference in the degree of analysis and not in the type. The current inconsistency in the Arab world between what is an index and what is a catalogue is likely to survive throughout the coming years.

To overcome the confusion in the nomination of the Arabic bibliographical products, this thesis considers relevant and falling inside its scope every product that meets the characteristics of printed indexes to periodical articles and uses the word index to describe such bibliographic tools.

In the Arab world there are different kinds of printed indexes. Some of these are the following:
2.2.1 Printed indexes to the contents of daily newspapers

These are indexes to the current affairs news published in dailies. They are not many in the Arab world, although currently there is a general trend towards the indexing of daily newspapers and weekly journals with the introduction of certain bibliographic retrieval systems which accept Arabic scripts. Unless leading Arabic daily newspapers take the lead in doing so, and establish indexes which cover the contents of all newspapers in each Arab country, there will be repetition of indexed news items on the country level. Some of the current and retrospective bibliographic tools which represent these indexes in the Arab world are the following:

a - The al-'Ahram Index (كشفائف الأهرام). It is a monthly index to the content of the Egyptian daily al-'Ahram (الأهرام). Al-Qadi (1974) mentioned in the introductory article of the first yearly cumulated printed issue of this index that the regular monthly issues of the al-'Ahram Index started to appear in January 1974.

b - The Yearly Analytical Index of 'Umm al-Qura Gazette (التحليلي السنوي لجريدة أم القرى). Badran and Tashkandi (1978) stated that this is a yearly index to the contents of the Saudi weekly official Gazette 'Umm al-Qura (أم القرى) and started publication in 1977.

2.2.2 Printed indexes to the contents of individual periodicals

This kind of index constitutes the largest section of indexes in the Arab world. Many periodicals produce volume by volume indexes to the contents of their several parts. However, only a few of them publish cumulative indexes to a number of volumes in order to make the task of retrospective searching for information easier. Such are al-'Arabi (العربي) which has one
that proceeds from 1958 to 1983, and *al-Darah* (الدارة) which has one that proceeds from 1975 to 1980. Both indexes have subject, author and title entries. Undoubtedly, the preparation of such indexes is an easy process in comparison to other available conventional indexes and its retrieval value is not as effective as in multiperiodical indexes. The individual indexes for Arabic periodicals currently appear in two forms:

**a - A printed yearly index which is more or less a combination of the lists of contents that appeared in the periodical throughout the year.** The index lists only the content lists of each separate issue of the periodical in one consecutive order according to the publication dates. These kinds of indexes are the most elementary guides to the contents of their periodicals. They are elementary because such a listing of contents pages is a page by page guide to the articles appearing in every single issue of the indexed periodical. Obviously this kind of index lacks essential features for the indexing process and should rather be called a collection of contents lists. One example of periodicals using this technique is the *Index of the Maghrebin History Journal* (كتشاف الجلة التاريخية المغربية) which was published in 1988 to list the contents pages of the Journal from issue number 1 to issue number 50 for the period between 1974 -1988.

**b - A periodical index which enjoys the characteristics of a proper index.** For instance it has the main parts which constitute an index to individual periodicals including author's name, article title, volume, part number, publication date, number of pages...etc. The periodicity of such indexes is often one year. Two examples of such Arabic indexes are the yearly cumulated printed indexes to the contents of the monthly periodicals *the Arab Future* (المستقبل العربي) and *al-Faysal* (الفيصل). Most of the periodicals which have individual subject indexing to their contents are
academic and many of them had their individual printed indexes done after quite a long time of their appearance or even after they had ceased publication. One of the oldest Arabic journals which had its cumulative individual printed index done after it had stopped publication is al-Muqtataff (المقتطف). The index was cumulated by Fu‘ad Sarruf and Linda Sadaqah of the American University of Beirut and was published in 1967-1968. It covered the contents of the journal since it first appeared in June 1876 until December 1952. ‘Abd al-Majid (1989) mentioned that the Arab Gulf States Information Documentation Centre has prepared individual retrospective printed indexes for nine current academic Arabic journals.

Although individual printed indexes for Arabic periodicals have recently begun to be published and are widely used, it is not an easy process to try to survey them accurately. For this, other bibliographic tools which deal with complete periodical items are needed. However, a limited survey for individual indexes for Gulf periodicals was tried here and out of 727 Arabic periodicals only 63 have individual printed indexes. These indexes are distributed among the Gulf States as shown in table 2.1.
<table>
<thead>
<tr>
<th>Country</th>
<th>Periodicals</th>
<th>Indexes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bahrain</td>
<td>44</td>
<td>3</td>
</tr>
<tr>
<td>Iraq</td>
<td>213</td>
<td>17</td>
</tr>
<tr>
<td>Kuwait</td>
<td>141</td>
<td>12</td>
</tr>
<tr>
<td>Oman</td>
<td>27</td>
<td>2</td>
</tr>
<tr>
<td>Qatar</td>
<td>31</td>
<td>4</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>185</td>
<td>19</td>
</tr>
<tr>
<td>United Arab Emirates</td>
<td>86</td>
<td>6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>727</strong></td>
<td><strong>63</strong></td>
</tr>
</tbody>
</table>

Table 2.1: Number of periodicals in each Arab Gulf State alongside the number of individual printed indexes that they publish. This table was prepared by using the Guide to Current Gulf Periodicals published in 1988 by the Arab Gulf States Information Documentation Centre.
Previously al-Sammadi (1983) stated that apart from the individual indexes which are included as a part of certain issues of periodicals at the end of each year, the number of indexing tools carried out in the Arab Gulf States for Arabic periodicals is 53. He provided a table which showed the different kinds of these tools and their geographical distribution. As far as the other Arab States are concerned it was not possible to construct similar tables from the available guides to their periodicals.

2.2.3 Printed indexes to the contents of a general group of periodicals

These are multidisciplinary indexes which analyse the contents of a group of different periodical titles by isolating the specific items of information within particular issues of periodicals both current and retrospective. Such printed indexes are usually approached by subject and in some cases by author name as well. The Arab world witnessed very few indexes in this category and most of them were short lived. Since this category forms the core of this thesis, it is going to be deeply studied and discussed in a separate section (see section 2.3).

2.2.4 Printed indexes to the contents of a specialized group of periodicals

Unlike the previous multidisciplinary ones, indexes of this kind are devoted to subject specializations. The proliferation of periodicals and the categorizations of subjects in sciences have led to the growth of such indexes. However, in the Arab world this kind of index, like most others, is not prosperous yet and the need for it has not yet emerged, especially with the scarcity of Arabic scientific periodicals. It is believed that the first Arabic specialized index in the Arab world was the Educational Index of Periodical Articles in the United Arab Republic (الكتاب التربوي لمقالات الدوريات).
which started in 1961 to cover the contents of 13 Arabic periodicals specializing in education, and was published by the Centre for Educational Documentation. Al-Hajrasi (1974b, p.35) mentioned this semi-annual index which ceased publication in 1967.

The other Arabic specialized printed indexes for retrospective literature in Arabic periodical articles were not many but they did cover quite a large period from the life of their chosen fields. Two retrospective specialized Arabic indexes prepared under the supervision of ALECSO (the Arab League for Education, Culture and Sciences Organization) are the following:

a - The Arab Subject Bibliography (البليوغرافيا الموضوعية العربية):

The idea behind this retrospective Arab Bibliography is to produce specialized bibliographies in different fields. It started in 1976 with the coverage of more than 60 different specialized Arabic periodicals which deal with Islamic subjects. 'Abu al-Nur (1978), (1980) mentioned that the first output of this project was six printed volumes of indexed articles which cover Islam in general and some related subjects and a seventh volume with authors' names arranged in alphabetical order. The Arab Subject Bibliography: Islamic Religion (البليوغرافيا الموضوعية العربية: علوم الدين الإسلامي) constitutes a classified subject bibliography according to the Islamic Religion Classification Scheme adopted by ALECSO, with an alphabetical subject index at the back of each volume.

The second output of this regional specialized Arabic Bibliography was, according to Khafaji (1980), the Arab Subject Bibliography: Education (البليوغرافيا الموضوعية العربية: التربية) which was published in 1980 by the Organization to cover the contents of more than 40 retrospective Arabic periodicals in the field of education. Entries in this bibliography are
arranged in alphabetical order according to authors' names and sometimes by article titles under standardized specific subject headings. It uses cross-references and provides two separate author and title indexes.

b - The Bibliographical Guide to the Arabic Writings on Libraries and Documentation (الدليل الببليوغرافي للإنتاج الفكري العربي في مجال المكتبات والتوثيق): this index covers what was published in Arabic in the fields of librarianship, documentation and information science inside and outside the Arab world during the twentieth century until 1976. The bibliographical entries were arranged alphabetically according to specialized subject headings. The index included bibliographic details on periodical articles, special reports, conferences and seminar proceedings as well as a few original and translated books. Afterwards, ALECSO published other bibliographical guides which cover the Arabic literature on librarianship and information science until 1985.

Recently the need was felt to cover the contents of the prolific non-scientific Arabic periodicals, this time by a private enterprise. Dilmun for publishing launched two current specialized printed indexes covering the contents of Arabic periodicals published inside and outside the Arab world. Bachir (1989) wrote about Dilmun's first specialized indexing product the Islamic Index (الكشف الإسلامي) which was first published in 1989 in Cyprus to be a quarterly current bibliographic index to the contents of 42 Arabic periodicals covering subjects related to the religion of Islam. The indexing technique followed in this index is different from other specialized Arabic printed indexes which use conventional methods and assign subject headings taken from prespecified authority lists. The publisher mentioned in the first issue that the index uses the permutation technique in displaying the keywords humanly selected from the texts of the indexed articles and then
sorted alphabetically by means of computer.

Auchterlonie (1990a) reviewed the index and stated that notwithstanding his criticisms, the index "is a very welcome addition to the all-too-select band of Arabic periodical indexes". The number of periodicals covered in this index seems to be enlarging. Farraj (1991) who reviewed the third issue of the index, stated that "the number of periodicals covered was 45". In the first issue it was 42.

Dilmun's second indexing product Palestine Record (الذاكرة الفلسطينية) published its zero issue in July 1992 to select and analyse the contents of 46 Arabic periodicals covering and discussing the Palestinian problem and the issues related to the Arab-Israeli conflict in the Middle East. The indexing technique followed in Palestine Record is similar to that used in Dilmun's first product the Islamic Index. Bachir (1992) reviewed the zero issue and mentioned that despite the shortcomings and deficiencies which exist in that issue, it covers a wide gap in the process of controlling the Arabic literature on the above matters and makes the retrieving process of required items much easier.

2.2.5 Printed indexes to the contents of collections of periodicals published by specific organizations

This category of printed index is produced by committees and academic institutions which prefer to pool their publications and get their contents analysed in one index. Generally, such indexes are retrospective and prepared by professionals. An early example of this kind of index is a sample index to the contents of the Arabic periodicals published in 1977 by UNESCO. It appeared in the issue number 21 of the bibliographical lists series produced by the Regional Committee of UNESCO in cooperation with the Egyptian
General Committee for the Book. The indexed journals are the following:

- *Future of Education* (مستقبل التربية)
- *UNESCO Message* ( رسالة الإنترنسيكو)
- *Science and Society* (العلم والمجتمع)
- *International Journal of Social Sciences* (المجلة الدولية للعلوم الاجتماعية)

Another recent retrospective index is the printed index to the contents of Arabic periodicals published by King 'Abd al-'Aziz University in Saudi Arabia. The *Analytical Index of King 'Abd al-'Aziz University Journals* (الكشف التحليلي لمجلات جامعة الملك عبد العزيز) was published late in 1992 by the King Fahd National Library. It covers the contents of all the periodicals published by the University from 1973 to 1983. In addition to the subject main entries, the index includes an author index and gives full bibliographic details of the indexed items.

### 2.2.6 Printed indexes to the contents of collections of periodicals published in specific countries

This category of printed indexes is rare in the Arab world and is mostly prepared for retrospective searching. In Iraq, the Centre for Scientific Documentation produces the *Analytical Index of Iraqi Journals* (الكشف التحليلي للمجلات العراقية) irregularly since 1979. In Kuwait the analytical retrospective index the *Kuwaiti Press in 25 Years* (الصحافة الكويتية في ربع قرن) covers what was written about Kuwait in 51 Kuwaiti newspapers and periodicals since December 1946 to September 1972. This index was compiled by M.H. 'Abd `Allah and published in 1974. In Saudi Arabia, a recent example of this category is the *Index of Saudi periodicals* (كشف الدوريات السعودية) which appeared as a pilot study in 1991 and was
published by the King Fahd National Library. Its introduction included statements clarifying the aims and the scopes of the index. Currently it covers the contents of some retrospective Saudi journals published by Saudi Universities. However, this is only a step towards covering all Saudi periodicals. Sharaf al-Din (1982) and al-'Annani (1989, p.163) respectively carried out studies regarding printed indexes which cover Saudi and Kuwaiti periodicals. This kind of index is as important as other kinds in controlling the Arabic literature published in Arabic periodicals.

2.3 Arabic indexes to the contents of a group of periodicals

In the previous sections, almost all kinds of Arabic printed indexes for periodical articles were approached. However, due to their reorganization according to other criteria, some of them will be mentioned in this section which is devoted to reviewing the state of the art of Arabic printed indexes to the contents of a selection of periodicals.

Indexes, in this section, are divided into two categories: retrospective which only includes indexes dealing with the contents of retrospective or non-up-to-date issues of a group of periodicals; and current which is issued periodically and analyses the contents of articles taken from various current periodical issues.

Whilst the availability of printed indexes to the contents of individual periodicals is essential and valuable, it is essential too that there should exist the means and tools of checking the contents through whole groups or selections of related periodicals cheaply and quickly. This function is fulfilled by a group of printed indexes which are usually alphabetical subject analyses of the article contents of a prespecified list of periodicals on a regular or irregular basis for current or retrospective periodical issues. Some of the
current regular indexes are cumulated into annual or larger volumes further to enhance searching efficiency. Such tools vary in the subject and coverage size, some are vast enterprises with no subject limitations and others are small and restricted to a certain specializations. In the Arabic language the largest multidiscipline current index is al-Fihrist (الفهرست) and the smallest specialized one is Palestine Record (الذاكرة الفلسطينية). Whereas the largest multidiscipline retrospective Arabic index is the Index Arabicus: 1876-1984 (كشف الدوريات العربية: 1876-1984) and the smallest one is the Subject Index of the Two Periodicals of the Institute of Public Administration (الكشف الموضوعي لجلتي معهد الإدارة العامة).

2.3.1 The Retrospective Arabic printed indexes

These are usually alphabetical subject analyses to the contents of given lists of retrospective issues of current or ceased publication periodicals. They vary in size and subject coverage and their subject indexing is generally done by professionals according to adopted rules with regard to subject headings assignment and bibliographic descriptions provided with their entries. By retrospective indexing is meant the subject indexing of back issues of live or dead periodicals.

In the Arab world, the number of retrospective printed indexes to Arabic periodical articles is larger than of those which index the contents of current periodicals. Due to the shortage of tools which provide full bibliographic descriptions on Arabic periodicals and the insufficiency of the available Arabic reference books, it was not possible to monitor the existing number of retrospective indexes for Arabic periodicals. However, it was possible at least to identify two major approaches to the organization of such retrospective indexing tools. In one approach, full bibliographic descriptions of articles
appear under humanly assigned subject headings arranged in alphabetical order. Article entries of this type are usually duplicated under more than one subject heading and cross-references are used both to link related subject entries, see also, and to provide alternative subject access used in the index, see. Most of such indexes include author indexes in separate sections. In the other major approach a form of classification is used. Bibliographic descriptions of articles are arranged under specific classification numbers or grouped under broad subject categories. For this indexing approach in the Arab world, subject indexes, although necessary, are not always provided to allow alternative subject approaches.

2.3.1. A Alphabetical subject indexes

One example of this approach is the recent retrospective index to the contents of Saudi periodicals specializing in library and information studies, Index of Articles in Saudi Periodicals Specialized in Libraries and Information (كشاف المقالات في الدوريات السعودية المتخصصة في المكتبات والعلومات). This index includes 378 entries to articles which previously appeared in three Saudi librarianship journals: Institute of Public Administration Library Bulletin (مكتبة الإدارة)، World of Books (عالم الكتب) and the Arabic Journal of Libraries and Information (مجلة المكتبات والعلومات العربية) since they were first published until the year 1989, except for the first which ceased publication in 1987. Article entries appeared under specific subject headings arranged alphabetically and at the back of the index another alphabetical author index was provided. One more feature is the extensive use of cross-references throughout the body of the subject index. A sample of this index is shown in figure 2.1.
التكشف

انظر أيضاً

التحليل الموضوعي

احمد محمد الشامية - أنظمة تكشيف واسترجاع الأروبه الصحري
- مكتبة الإدارة،  مج. 11،  ع 2،  جمادي الأول،
  1404 هـ/ فبراير 1984 م، ص ص 72 - 104.
  (125)

 أهمية مصطلح صادق - الأسس والملامح الرئيسية للكانز العلوم
 الاجتماعية - مجلة المكتبات والمعلومات العربية،
 ع 4، محرم 1406/ أكتوبر 1985 م، ص ص 77- 90.
  (126)

 حشمت قاسم - كشف الكلمات المفتاحية في السياق - عالم
 الكتب،  مج. 5،  ع 4،  ربيع الآخر 1405 هـ/ يناير 1986 م
  ص ص 128 - 150.
  (127)

 عايدة ابراهيم نصير - التكشيف المبكر في العالم العربي ومفترقات
 التكشيف الحالي - مجلة المكتبات والمعلومات العربية،
 س 9،  ع 2،  نو. الحجة 1409 هـ/ يوليو 1989 م، ص ص
  152 - 185.
  (128)

Figure 2.1 : sample of the Index of Articles in Saudi Periodicals specialized in Libraries
and Information
Various characteristics of this index are worth attention:

a - Specific subject headings are used.

b - Full bibliographic details are provided for each entry.

c - Bibliographic descriptions are duplicated under more than one entry, i.e. more than one subject heading is assigned to each article.

d - Cross-references *see* and *see also* are often used throughout the subject index.

e - Provides a separate, but not self-contained author index to the associated subject index.

Other Arabic printed indexes which more or less enjoy the same characteristics are briefly mentioned hereafter according to their publication dates:

- **Index of Saudi Periodicals** (كتشاف الدوريات السعودية). Published by King Fahd National Library in 1991 as a pilot study.


- **Index of Arabic Periodicals at the Library of the Institute of Public Administration** (كتشاف الدوريات العربية في مكتبة معهد الإدارة العامة). Published by the Institute in 1988.


- The **Subject Index to of the Two Periodicals of the Institute of Public Administration** (الكتشاف الموضوعي لمجلتي معهد الإدارة العامة).
- Researches and Articles of Arabic Educational Journals at the Library of Educational Documents: Subject Index (أبحاث ومقالات مجموعة الجلات العربية الترجمة بمكتبة الوثائق التربوية: كشف موضوعي). Published by the Centre for Statistical Data and Educational Documentation in Saudi Arabia, first in 1978 then in 1982.

- The Analytical Bibliography of Articles published in the Arabic Periodicals at the Periodicals Library of the Faculty of Commerce in Kuwait University (الباليغرافيا التحليلية للمقالات الواردة في الدوريات العربية بمكتبة الدوريات بكلية التجارة – جامعة الكويت). Published by the University in 1977 and 1979.


2.3.1.B Classified indexes

From amongst the totality of retrospective Arabic printed indexes, two types of retrospective classified Arabic printed indexes were seen. In the first, article entries appear under highly specific classification numbers usually derived from a general or special purpose Arabic classification schemes. An example of this type is the Arab Subject Bibliography: Islamic Religion (الباليغرافيا الموضوعية العربية: علوم الدين الإسلامي). This Bibliography which was published by ALECSO started in 1976 and stopped in 1978 after the publication of six printed volumes of bibliographic details about subjects related to Islam in general, Qur’anic studies, prophetic tradition, life of the prophet Muhammad, Islamic jurisprudence and theology. Bibliographic citations in this bibliography are arranged according to the Islamic Religion Classification Scheme (تصنيف علوم الدين الإسلامي) and
the class number of each record is shown alongside its subject heading. For further details see section 2.2.4 of this chapter.

In the second type, article entries appear under broad subject categories with specific subdivisions. An example is Index Arabicus : 1876-1984 (كشاف الدوريات العربية : 1876-1984). It is a retrospective printed classified index of articles relating to Arabic and Islamic studies published in Arabic scholarly periodicals. It was compiled by ‘Abd al-Rahman (1989) and published in Baghdad by the Arab Gulf States Information Documentation Centre. It analyses in four volumes, the contents of about 250 periodical articles as well as items from about 50 collective works and conference proceedings. The compiler of this index has chosen to index Arabic periodicals which appear no more than once a month except for the weeklies the Culture (الثقافة) and the Message (الرسالة).

Auchterlonie (1991) reviewed the index and said that "the internal organization and layout, the Kashshaf uses the same pattern as Index Islamicus, although most subject headings have been considerably expanded". The index considers the Index Islamicus as its distant parent and follow the standard ISBD (International Standard Bibliographic Description) pattern of author, article title, journal title, issue number, date and pagination to provide the articles bibliographic details. A sample of this index is shown in figure 2.2.
المعاجم ومجموعة المصطلحات الحديثة
مرتبة حسب الموضوعات

الكتب (مايو - يونوو 1961)
صف 33 - ص 35

شمع خليفة

(أ) - المكتبات

(ب) - الفلسفة

تبرير شيخ الأرض

ط - مجموعات علم الكتب : عالم الكتب 1667 (مايو - يونوو) ص 3 - ص 18

(ب) - مجموعات علم الكتب : عالم الكتب 1967 (يناير - فبراير) ص 24 - ص 38

مصطلح بنمرس

- مجموعات الأعلامية : عالم الكتب 1972 ص 219 - ص 237

- ميلم : مجموعات العلمية : عالم الكتب 1979 ص 130 - ص 138

- يوصف اسم داير

- مجمع المصطلحات الكبيرة / فوزي راجي ابراهيم وفضلا رزق عمال

Figure 2.2 : sample of Index Arabicus
2.3.2 The current Arabic printed indexes

Whilst retrospective indexing is relevant and essential for specialists to obtain as much information as can be found regardless of date, current indexing is meant to keep them up-to-date with their area of specializations.

Arab specialists and authors interested in Arabic librarianship and information studies agree that the first multiperiodical Arabic printed index to current Arabic journals appeared in Egypt in 1962, that is almost fifteen years after the 1948 call from the Culture Committee in the Arab League for the production of current indexes for Arabic literature.

In the Arab world, seven trials for producing multiperiodical current printed indexes to the contents of Arabic periodicals were traced in the literature on Arabic librarianship and information studies. The outputs of these trials will be divided into two parts. The first will review with brief details the indexes which officially ceased publication due to certain reasons. The second part will be concerned with reviewing those indexes which still appear on a regular or irregular basis, no matter whether they are produced by private enterprises or official bodies.

2.3.2.A Ceased publication indexes

Two Arabic multiperiodical indexes are included in this category. These two indexes were short-lived and Egyptian oriented as regards their coverage. However, one of them included sources from other Arab countries in its latest issues. Although, both of them were meant to be current indexes, they did not have much in common. The first one the Analytical index of Arabic Newspapers and Journals (الكشف التحليلي للصحف والجلات العربية) was general purpose and published on a monthly basis, whereas the other the
Educational Index of Periodical Articles in the United Arab Republic
الكشف الترقي لمقالات الدوريات في الجمهورية العربية المتحدة
was a specialized index and published every six months. Three common things between these two indexing tools can be mentioned after they ceased publication:

a - The choice of periodicals for indexing showed a bias towards Egypt.

b - They were short-lived indexes.

c - They published their first issues in the same year in 1962.

In the following two sections, each of these two indexes will be discussed briefly in an attempt to indicate some of the evident reasons behind their early demise.

2.3.2.A.1 The Analytical Index of Arabic Newspapers and Journals

Although this index appeared with another index in the same year in 1962, Arab specialists always repeat that it was the first Arabic index which analysed the contents of several Arabic periodicals. Al-Hajrasi (1974b, p.37) mentioned that this index was compiled by a group of librarians who after their graduation from Cairo University planned to launch such a tool. The first issue of the index was published in January 1962 as a monthly index to cover, according to al-Shinniti (1962), the contents of about fifty periodicals. However, the issue number eight of this index which was published in August 1962 covered the contents of thirty nine periodicals consisted of five newspapers, twelve general journals and twenty two specialized periodicals. In indexing the selected items, the indexers followed modern and standardized rules and gave full bibliographic details in each bibliographic record. Al-Hajrasi (1974b, p.37) stated that the index used to appear with minor or major changes in its subject headings, cross-references and in the
number of selected periodicals in every single issue and that the unresolved problem was the constant delay in its appearance until it ceased in 1967. The first issues of the index consisted of two sections. One was for authors and subjects in one alphabetical sequence and the second was for news and articles about laws and governmental decisions. However, in the last issues it was divided into three sections: subject index, author index and laws and governmental decisions index. The index was planned to be cumulated annually, but never appeared in such cumulation. A sample page of this index is displayed in figure 2.3.

2.3.2.A.2 The Educational Index of Periodical Articles in the United Arab Republic

This index was first published in April 1962 to cover the period ranging from January to June 1961. 'Abd al-Hadi (1982, p.181) said that the index was launched to be a current biannual specialized index which analyses the contents of 13 Egyptian periodicals specializing in Education. However, it appeared in its latest issues to be covering some other non-Egyptian periodicals on an annual basis. Article entries in this index appeared under alphabetical subject headings and contained details such as article title, author name, source name, volume and issue numbers, publication date and number of pages. It ceased publication at the end of 1967. A sample page of this index is displayed in figure 2.4.

Although the above two indexes were in all respects only exploratory trials in the Arab world, they carried inside them right from the very beginning indications of their short-living period. Al-Hajrasi (1974b, p.37) mentioned that the Analytical Index of Arabic Newspapers and Journals used to be compiled more or less by volunteer indexers who started to disappear
one after the other until the index finally stopped processing. In addition, the size of the index grew gradually due to the inclusion of newspaper items, which led to more expense and then greater loss. The other tool, the Educational Index of Periodical Articles in the United Arab Republic, had the problem of finance and the unwise planning as regards to coverage and periodicity.
Figure 2.3: sample of the Analytical Index of Arabic Newspapers and Journals
وجهة نظر في اصلاح التعليم

الدبيبي، النشرة العربية

البحرين، أ.ف. (1975) ص

11 1684

طريقة

التعليم بالقراءة - معالج اللغة

طرق التعليم - أمريكا

أجر أداء النباس باللغة العربية

في الإمارات، رحلة موسية

1889-11-1684

سعودية، أ.ف. (1975) ص

11 1684

 чтобы نظر في اصلاح التعليم

الدبيبي، النشرة العربية

البحرين، أ.ف. (1975) ص

11 1684

طريقة

التعليم بالقراءة - معالج اللغة

طريقة التعليم - أمريكا

وجهة نظر في اصلاح التعليم

الدبيبي، النشرة العربية

البحرين، أ.ف. (1975) ص

11 1684

Figure 2.4: sample of the Educational Index of Periodical Articles in the United Arabic Republic
2.3.2.B Current publication indexes

Five of the seven current Arabic indexing tools will be reported under this section. Some of them may have stopped publication, but since none of the responsible committees or bodies has officially declared this about any of them, all indexes are going to be treated as current tools.

Three of these current indexes belong to the approach of alphabetical subject headings arrangement. These are the Analytical Index of Iraqi journals (الكتشاف التحليلي للمجلات العراقية), al-Fihrist (الفهرست) and the Analytical Index of the Arabic and Foreign Periodicals and Bulletins Available at the Information and Documentation Centre of the Gulf Cooperation Council Secretariat General (الكتشاف التحليلي للدوريات والنشرات العربية والأجنبيّة المتوفّرة في مركز التوثيق والمعلومات بالأمانة العامة لمجلس التعاون لدول الخليج العربي).

Lancaster (1991, p.138) described indexes which follow this approach as alphabetico-specific indexes. In this indexing approach the alphabetical subject index contains subject entries and cross-references arranged alphabetically in one sequence and main entries are selected from prespecified general or specialized subject headings lists. The subject entries consist of subject headings or keywords and bibliographic details which identify the item being indexed. Cross-references are often of two kinds: see references, which as described by Foskett (1982, p.281) "lead from headings which appear only in the entry vocabulary (i.e. terms which users may think of but which are not part of the preferred index vocabulary) to headings in the index vocabulary; and see also references, which serve to link related headings within the preferred index vocabulary, either to show semantic relationships or to reveal 'hidden' terms in a composite heading".
The remaining two current Arabic printed indexes also belong to the alphabetical subject headings approach. However, they follow a different indexing technique. According to the publisher of both the *Islamic Index* (الكشف الإسلامي) and the *Palestine Record* (الذاكرة الفلسطينية) the index to the article is "an alphabetical permuted index of keywords followed by a number (opposite to the main entry or its minor entry) referring to the reference number which precedes the specific record listed in the article section of the bibliographical citations". It is obvious that the publisher meant to describe his two products as permuted indexes. However, he was like most others confused between the permutation and the rotation techniques and really they were rotated indexes. Neither tool includes any kind of cross-references in their indexes to articles.

In the following five sections, each of the above mentioned indexing trials is going to be briefly discussed. However, more emphasis will be placed on the *al-Fihrist* trial, since it has been the leading indexing tool in the Arab world despite all its accompanying drawbacks. Emphasis will also be placed on the latest two indexing trials in an attempt to see whether they could overcome any of the hurdles which are currently facing al-Fihrist.

2.3.2.B.1 The Analytical Index of Iraqi Journals

This is an analytical index to the contents of Iraqi Journals. The number of journals which are covered by this tool varies from one issue to another and according to al-Hazimi (1990, p.396), this indexing tool has been published by the Centre for Scientific Documentation in the Iraqi Council of Ministers since 1979 on an irregular basis. Article entries in this index are arranged alphabetically under the selected subject headings. No studies were found in Arabic or any other language concerned with the structure and
efficiency of this index. However, it seems to be a conventional pre-coordinated printed index depending on controlled subject headings taken from prespecified subject heading lists. Also, its irregularity reveals that it is not a promising tool for current Arabic literature.

2.3.2.B.2 Al-Fihrist

With the publication of Al-Fihrist, which means the index in Arabic, it could be said that a proper indexing tool to the contents of Arabic periodicals has seen the light. Despite the fact that a few trials to produce Arabic printed indexes for Arabic periodicals were carried out before its appearance, Al-Fihrist is widely considered as the first effort of its kind in the Arab world. Ali (1988) described it as the noteworthy index of all indexing tools in the Arab world. Al-Fihrist is a quarterly index published in Lebanon by a commercial enterprise called Al-Fihrist Academic Research Institute.

The first issue of Al-Fihrist appeared late in 1981 to cover the contents of 66 Arabic periodicals ranging from monthlies to quarterlies and was published during the first three months of the same year. Before, Al-Fihrist appeared in two zero issues. The first was in March 1981 and the second was in June of the same year. Regarding the arrangement, each issue of Al-Fihrist contains two separate indexes. The first is a subject index and the second is an author index. Article entries in the subject index are arranged alphabetically by author under Arabic subject headings taken from three prescribed subject heading lists. Two of these are used by the American University of Beirut Library and by Al-Nahar newspaper Research Centre. The third list is the Arabic Subject Headings prepared by the Library Cataloguing Department at University of Riyadh in 1978. The citation for each entry consists of author, article title, source name, volume and issue number as well
as number of pages and publication notes such as bibliographies, images, graphs and so on.

As far as the author index is concerned, it presents full bibliographic citations of author signed articles arranged alphabetically by article title under authors names. See and see also cross-references are heavily used throughout the subject index. In issue number 23/24 eighteen see also references appeared under the main term arms and 56 under the term Islam.

Al-Fihrist is a multidiscipline index. It covers almost all disciplines including administration, arts, economics, education, geography, history, law, philosophy, sociology and to a very small and limited extent subjects in science and technology. In fact al-Fihrist is a non-scientific indexing tool rather than a general one and this limitation relates to the indexing policy and to the scarcity of Arabic Language scientific periodicals. Ali (1987) mentioned that the majority of periodicals covered by al-Fihrist is published by academic institutions, societies and various publishing houses in and outside the Middle East. With regard to the number of the periodicals selected for indexing, this has been variable to some extent throughout the life of the index. The first issue, which covered periodicals published in the first quarter of 1981, indexed the contents of 66 periodicals, then the number increased to 104 in the fourth issue of the index. After that, the number of indexed periodicals diminished to 62 in the issue number five. Afterwards the index seems to have been through some difficulties. Thus, the subsequent issues 6 and 7 appeared in one volume covering the contents of 82 periodical articles. For the following issues number 8 which covered 114 periodicals and number 9 which covered 107 periodicals, the index returned to its original policy of one issue in one volume. However, for the remaining issues, the policy of producing two issues in one volume prevailed but
exceptionally not for the issue number 12 which appeared in one single volume. The latest volume of al-Fihrist which appeared in April 1993 included issues 31/32 and covered the contents of 81 periodical articles published in the years between 1985 to 1988. Ubaidly (1988a) stated that "al-Fihrist is not committed to periodicals which must be treated in each issue. Every issue adds new editions of periodicals which were not indexed in earlier issues. This confuses the user and deprives him of the chance to practise organized and continuous searching ". With regard to the total number of indexed periodicals in al-Fihrist, al-'Atram (1989) mentioned that although it is not easy to give an exact number due to the inconsistency in coverage, it can be said that the overall number is about 160 periodicals per issue. In comparison with the growing number of Arabic specialized and academic periodicals, this number seems to be relatively small.

Since April 1993, no more issues of al-Fihrist have been produced. These arrears in producing new issues of al-Fihrist do not seem to be normal and it rather seems to be a cessation of publication. But since no official statement has been issued or announced, the index is considered here as a current indexing tool.

Whether this cessation is only a mere delay or an actual cessation of the publication itself, this constitutes bad reputation for the leading indexing tool in the Arab world. As a matter of fact, one of the main drawbacks of al-Fihrist has been its high price from the very beginning of its publication. This high price is probably a reflection of the indexing process and the publishing costs. In other words, making the yearly subscription rate to al-Fihrist higher than normal may have been imposed by the costs involved in the conventional production of this index. Al-Fihrist is a conventional index in its indexing method and publishing criteria. The indexing process is carried out manually
by professional indexers, the sorting process is also carried out manually. These two processes are time consuming and costly. Added to this are the virtues of printing and typography which are of higher standard than normal, which in turn add more costs to the production of the index.

Auchterlonie (1982) who reviewed the first issue of al-Fihrist criticized its high price and stated that "given its (al-Fihrist) disadvantages, the number of its subscribers may prove too low to make it a commercial proposition ". Al-Shaykh (1983), who was assassinated in Lebanon late in 1985 while editor in-chief for al-Fihrist, mentioned that the number of subscribers to al-Fihrist does not exceed 300. Admitting the high price of al-Fihrist and the criticisms of specialists, al-Shaykh (1985) stated in an interview with a weekly Lebanese Journal that with the absence of official and governmental support, it was not possible at all to make the subscription rate less than 200 American Dollars at that time. It is believed that al-Fihrist has been facing financial difficulties soon after its first few issues and that only private subsidies had made it last longer.

Despite the fact that the publication of al-Fihrist does fill a serious gap in existing Arabic bibliography, yet it is not a perfect tool to index the contents of Arabic periodical articles. Apart from the above mentioned drawbacks, specialists have also criticized the structure of some of its subject headings. Talib (1982) wrote an extensive critical study for al-Fihrist and stated some of the disadvantages caused by the way al-Fihrist's indexers structure and assign the subject headings to the indexed articles. In his study he called for al-Fihrist's indexers to overcome such disadvantages so that the index can survive and fulfill its job. Long after this critical review, al-'Atram (1989) wrote a new critical study which concluded that al-Fihrist index has to cover more periodicals and pay attention to the structure of subject headings.
within the subject index. He also suggested that a computer should be used to avoid delay and to produce cumulations. Finally, al-'Atarm (1989) proposed for al-Fihrist to merge both the subject index section and the author index section into one alphabetical order. It is relevant to mention that according to Talib (1982) this suggestion was tried before and al-Fihrist appeared merged in a pilot index in 1981 and was distributed to a limited number of specialists. However, the reaction for this sorting was not published.

Gorman and Mills (1992, p.97) assessed the content of al-Fihrist and although they cited some incorrect information, their conclusion made sense in that al-Fihrist "is not a suitable resource for the most current information on the Middle East - although in many fields it is the only guide to literature in Arabic". A sample page of al-Fihrist is shown in figure 2.5.

2.3.2.B.3 The Analytical Index of the Arabic and Foreign Periodicals and Bulletins Available at the Information and Documentation Centre of the Gulf Cooperation Council Secretariat General.

As indicated in its title, this indexing tool is produced by the Secretariat General of the Gulf Cooperation Council. Al-Shatti (1991) mentioned that the index is published quarterly and distributed to governmental institutions and committees in the Council States as well as to researchers and specialists interested in the area. It covers the contents of about 100 publications consisting of periodicals, reports and bulletins in the fields of politics, economics, law, information studies, finance and management, military affairs, media and environment. The Information and Documentation Centre in the General Secretariat started this indexing process in 1982 and in 1987 STAIRS was introduced as a bibliographic retrieval system. However, so far no annual index has been published. Article entries for this index are
arranged under alphabetically sorted subject headings taken from the Arabic Subject Heading List prepared by al-Khazindar.

Ubaidly (1987) evaluated the first issue of the fourth year which covered the contents of the available Arabic and foreign periodicals and bulletins in the Centre for the period between January and March 1985. He observed that the subject index lacks the required distinction and consistency in assigning specific and general terms. Some general articles were given specific subject headings and some other specialized articles were put under general terms. In fact this index is not promising and has many drawbacks: for example it has no see cross-references and only one subject heading was assigned to each article.

With regard to the bibliographic descriptions of the indexed items, the index does not mention any specific rules used, although most of the required details are indicated.

This index is not as known as al-Fihrist in the Arab world and its usage and distribution are limited mainly to governmental bodies and specialists in the Gulf States affairs. In addition to that the index only covers the contents of the available periodicals, bulletins and reports which are forwarded to the Centre either through subscriptions or through gifts. Al-Shatti (1991) stated that the index concentrates on publications dealing with Gulf States issues.
عدة التغيرات السياسية الحديثة في النظم العربية (1988)

ال_sampler

الثقافة العربية، نشر إصدار الديون العربية بمدارك الديون العربية، 1988/2، ص 129، 131، 139، 141، 143، 145، 147، 149، 151، 153، 155، 157، 159، 161، 163، 165، 167، 169، 171، 173، 175، 177، 179، 181، 183، 185، 187، 189، 191، 193، 195، 197، 199، 201، 203، 205، 207، 209، 211، 213، 215، 217، 219، 221، 223، 225، 227، 229، 231، 233، 235، 237، 239، 241، 243، 245، 247، 249، 251، 253، 255، 257، 259، 261، 263، 265، 267، 269، 271، 273، 275، 277، 279، 281، 283، 285، 287، 289، 291، 293، 295، 297، 299، 301، 303، 305، 307، 309، 311، 313، 315، 317، 319، 321، 323، 325، 327، 329، 331، 333، 335، 337، 339، 341، 343، 345، 347، 349، 351، 353، 355، 357، 359، 361، 363، 365، 367، 369، 371، 373، 375، 377، 379، 381، 383، 385، 387، 389، 391، 393، 395، 397، 399، 401، 403، 405، 407، 409، 411، 413، 415، 417، 419، 421، 423، 425، 427، 429، 431، 433، 435، 437، 439، 441، 443، 445، 447، 449، 451، 453، 455، 457، 459، 461، 463، 465، 467، 469، 471، 473، 475، 477، 479، 481، 483، 485، 487، 489، 491، 493، 495، 497، 499، 501، 503، 505، 507، 509، 511، 513، 515، 517، 519، 521، 523، 525، 527، 529، 531، 533، 535، 537، 539، 541، 543، 545، 547، 549، 551، 553، 555، 557، 559، 561، 563، 565، 567، 569، 571، 573، 575، 577، 579، 581، 583، 585، 587، 589، 591، 593، 595، 597، 599، 601، 603، 605، 607، 609، 611، 613، 615، 617، 619، 621، 623، 625، 627، 629، 631، 633، 635، 637، 639، 641، 643، 645، 647، 649، 651، 653، 655، 657، 659، 661، 663، 665، 667، 669، 671، 673، 675، 677، 679، 681، 683، 685، 687، 689، 691، 693، 695، 697، 699، 701، 703، 705، 707، 709، 711، 713، 715، 717، 719، 721، 723، 725، 727، 729، 731، 733، 735، 737، 739، 741، 743، 745، 747، 749، 751، 753، 755، 757، 759، 761، 763، 765، 767، 769، 771، 773، 775، 777، 779، 781، 783، 785، 787، 789，791، 793، 795، 797، 799، 801، 803، 805، 807، 809، 811، 813، 815، 817، 819، 821، 823، 825، 827، 829، 831، 833، 835، 837، 839، 841، 843، 845، 847، 849، 851، 853، 855، 857، 859، 861، 863، 865، 867، 869، 871، 873، 875، 877، 879، 881، 883، 885، 887، 889، 891، 893، 895، 897، 899، 901، 903، 905، 907، 909، 911، 913، 915، 917، 919، 921، 923، 925، 927، 929، 931، 933، 935، 937، 939، 941، 943، 945، 947، 949، 951، 953، 955، 957، 959، 961، 963، 965، 967، 969، 971، 973، 975، 977، 979، 981، 983، 985، 987، 989، 991، 993، 995، 997، 999، 1001، 1003، 1005، 1007، 1009، 1011، 1013، 1015، 1017، 1019، 1021، 1023، 1025، 1027، 1029، 1031، 1033، 1035، 1037، 1039، 1041، 1043، 1045، 1047، 1049، 1051، 1053، 1055، 1057، 1059، 1061، 1063، 1065، 1067، 1069، 1071، 1073، 1075، 1077، 1079، 1081، 1083، 1085، 1087، 1089، 1091، 1093، 1095، 1097، 1099، 1101، 1103، 1105، 1107، 1109، 1111، 1113، 1115، 1117، 1119، 1121، 1123، 1125، 1127، 1129، 1131، 1133، 1135، 1137، 1139، 1141، 1143، 1145، 1147، 1149، 1151، 1153، 1155، 1157، 1159، 1161، 1163، 1165، 1167، 1169، 1171، 1173، 1175، 1177، 1179، 1181، 1183، 1185، 1187، 1189，71

Figure 2.5 : sample of the al-Fihrist
2.3.2.B.4 The Islamic Index

Bearing in mind the drawbacks of al-Fihrist and the need for better indexing tools, Ubaidly Ubaidly, who used to be the editor in chief of al-Fihrist, left its editorship and moved to Cyprus to establish an independent private publishing company called "Dilmun for Publishing". The Islamic Index was the first indexing product of this new publishing house. In fact it is the Arabic equivalent to Index Islamicus which covers Western language publications. As mentioned earlier, the Islamic Index was first published in 1989 as a quarterly index to cover the contents of Arabic periodicals specializing in the religion of Islam.

For the Islamic Index to be a better indexing service, Ubaidly decided to use the facilities provided by computers. Moreover, he gave up the idea of conventional indexing and relied on a different indexing technique which he called, in the first issue of the index, the permutation technique. As mentioned earlier, the publisher was mistaken and really the technique was rotation.

Although the rotation technique is used and computer facilities are involved, the Islamic Index is not generated automatically and not yet published in any cumulated form. In a private conversation, Ubaidly (1992) stated that the keywords are humanly derived from the full text of each article selected for indexing and from other Arabic subject heading lists such as the thesaurus of al-Jami'a (الجامعة). They are then fed into the computer system which carries out the sorting process according to prescribed criteria by rotating them into two indexing positions which are called "main entry" and "minor entry". Other keywords which belong to the same document are listed in alphabetical order under the minor entry. However, it is observed
that minor entries which in all records follow the main entries are neither listed alphabetically nor according to other sorting criteria. To distinguish between these two kind of " entries " or terms, the editor preferred to print the main entry in bold type. Alongside each minor entry, a numerical reference code is provided to refer users to the article's citations in the index.

The number 20 of this Index covers the contents of 59 periodicals ranging from monthlies to annuals for the period between October and December 1992. After reviewing the first issue of this index, Auchterlonie (1990a) stated that " there is a lot of leeway to be made up before al-Kashshaf (Islamic Index) can be considered as wide-ranging and thorough an index for Arabic-language publications as Quarterly Index Islamicus / Index Islamicus is for Western-language material, but at least it is a start " . He also mentioned that " if the editor is able to refine the terms of reference and to expand the coverage, al-Kashshaf will become as indispensable for Arabic-language material as Index Islamicus is for Western-language studies ".

Farraj (1991), who reviewed issue number 3 of volume 1, questioned the comprehensiveness in the number of periodicals selected for indexing and mentioned that the reviewed issue of the Islamic Index covers the contents of 33 periodicals specializing in Islam, 8 specializing in other non-scientific subjects and another 4 general periodicals. He believes that for this index to be thorough and comprehensive, the editor should increase the number of selected periodicals for indexing. Farraj himself (1991) found out in a previous study that, apart from other non-Arab Islamic countries, the number of current Arabic periodicals specializing in Islamic studies exceeded 100.

In addition to this number of specialized periodicals in Islamic studies, it is relevant to mention that almost all non-scientific periodicals in the Arab world touch on in one way or another studies related to Islam. After all, as
Roper (1986, p.22) who is editing *Index Islamicus* stated, "Islam embraces all spheres of human life, not just religion in the narrow, Western Christian sense of that term" as is thought by some people. In his turn, Auchterlonie (1990a) said that "it is not always easy to draw a distinction between Islam as a religion and Islam as a culture." The purpose of quoting the above specialists is to reveal the fact that the periodicals' coverage of the *Islamic Index* is inadequate with regard to the available large number of current Arabic periodicals which deal with Islamic studies.

It is believed that the use of computer facilities to produce the *Islamic Index* has been to some extent successful in making the index more current and thus the witnessed delay period in *al-Fihrist* between the publication of a periodical issue and its appearance in the index has been minimized. However, the *Islamic Index* has not been successful in covering enough publications. This is due to the fact that, although rotation has been chosen by the editor as the indexing technique, indexers still have to go through the whole articles to select and sometimes assign representative keywords for the indexed articles. In other words, this means that if the editor decided to cover more periodicals this would compel him to either increase the number of indexers or to face inevitable delay in producing the index. It is relevant to mention here that the printing of the *Islamic Index* is financially supported by the 'Imam 'Uza'i College of Islamic Studies. A sample of this index is shown in figure 2.6.
<table>
<thead>
<tr>
<th>الكتاب الإسلامي</th>
<th>المحتوى</th>
</tr>
</thead>
<tbody>
<tr>
<td>الديانة الحكيمة</td>
<td>315 تعليم الآخرين</td>
</tr>
<tr>
<td>حفظ وبيانات</td>
<td></td>
</tr>
<tr>
<td>الجزء الثاني:</td>
<td></td>
</tr>
<tr>
<td>الإسلام</td>
<td>31 شهرزاد</td>
</tr>
<tr>
<td>القدرة الإسلامية</td>
<td>31</td>
</tr>
<tr>
<td>القلي</td>
<td></td>
</tr>
<tr>
<td>علوم في الشريعة</td>
<td></td>
</tr>
<tr>
<td>السيرة النبوية</td>
<td>66</td>
</tr>
<tr>
<td>الحكيم</td>
<td></td>
</tr>
<tr>
<td>سلامة ومؤثرات</td>
<td></td>
</tr>
<tr>
<td>الخلافة الوليدية</td>
<td></td>
</tr>
<tr>
<td>الاسلامية</td>
<td>36</td>
</tr>
<tr>
<td>الإمارات</td>
<td></td>
</tr>
<tr>
<td>الولايات المتحدة</td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.3.2.B.5 The Palestine Record

This index is the second output of Dilmun for Publishing. Its first issue was published in Cyprus in late 1992. Bachir's (1992) review to its zero issue which was launched in July 1992 revealed that there is a lot of similarity with the Islamic Index as regard to the use of computer facilities and to the use of the rotation technique in indexing the contents of the selected periodicals. As well, the small number of indexed periodicals would threaten its continuation especially with the proliferation of Arabic periodicals which deal in one way or another with the speciality of this index, which was declared in its zero issue.

The editor mentioned that the Palestine Record is a quarterly index produced to cover the contents of publications dealing with the Palestinian problem and the conflict between Israel and the Arab States. It is needless to say that the issue of Palestine and the Arab-Israeli conflict has influenced every political, economic and social aspect of life in the Arab world. Thus, almost every non-scientific Arabic periodical has approached this topic in some way.

The zero issue was very weak in its coverage. It indexed the contents of 46 periodicals ranging from monthlies to quarterlies. It is believed that the obstacles facing the progress of this index are the same of those facing the Islamic Index. On one hand, increasing the number of the selected periodicals for indexing would involve an increase in the number of indexers and therefore would add unbearable additional costs to the production process. On the other hand leaving the index unchanged would risk its continuation as a thorough and useful indexing tool.
Finally, one more common feature between Islamic Index and Palestine Record is that each of them has designated separate subject index and bibliographic description sections for books related to the topic of their interests. Whether these sections are sufficient or not is beyond the scope of this thesis. However, it seems that sections related to periodical articles have taken more effort, and that much more attention has been given to them by Dilmun's for Publishing indexers. A sample of Palestine Record is shown in figure 2.7.
<table>
<thead>
<tr>
<th>Number</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>الأردن</td>
</tr>
<tr>
<td>2</td>
<td>فلسطين</td>
</tr>
<tr>
<td>3</td>
<td>إسرائيل</td>
</tr>
<tr>
<td>4</td>
<td>قضاء النوارس</td>
</tr>
<tr>
<td>5</td>
<td>الديوان</td>
</tr>
</tbody>
</table>

Figure 2.7 : sample of the Palestine Record
2.4 Other printed bibliographical services

In addition to the independent or stand alone indexing tools for the contents of Arabic articles, there are some other printed bibliographical services which include analysis of the contents of some current Arabic periodical articles. These bibliographies appear in special sections inside a few Arabic periodicals as an additional service to their readers and usually cover monographs as well.

It is believed that the first bibliography of such services appeared in 1968 right after the cessation of the Analytical Index of Arabic Newspapers and Journals. Al-Hajrasi (1974b, p.38) mentioned that the cessation of the only current index to the contents of Arabic periodicals encouraged 'Abu al-Futuh 'Awdah to publish in 1968 in issue number 42 of the Egyptian quarterly the Arabic Book (الكتاب العربي) what he called the Scientific Researches of Journals in the United Arab Republic (بحوث مجلات ج. ع. م.) which indexed the contents of about 40 Arabic and foreign journals taken from the previous year 1967. 'Awdah continued to publish his yearly bibliography during July of the following three years 1969, 1970 and 1971 to cover respectively the contents of periodicals published in 1968, 1969 and 1970. According to al-Hajrasi, 'Awdah hoped to publish in the issue number 58 of the Arabic Book a five year cumulated index which was estimated to include about 5000 main entries in addition to 1500 main entries for the year 1971, but the Arabic Book ceased publication in 1971 after the issue number 54 which included the index for the contents of 1970 journals.

As a matter of fact, such bibliographic services cannot be considered as reliable and consistent indexing tools for many reasons:

a - As noticed earlier, the fate of such tools is attached to the fate of the
periodicals which include them. Sometimes the cessation of a certain periodical may have nothing to do with the quality of the included bibliography.

b - Usually such indexing tools are prepared by individuals, thus they are bound for delay in indexing current periodical issues.

c - Usually in such bibliographic services, article entries are arranged under broad subject headings and only one heading is assigned to each article.

d - Almost all Arabic periodicals which include such indexing tools provide them as complementary services to their readers, thus they are not paid much attention as regards to consistency in covering consecutive issues of certain journals.

These reasons are more than enough for this kind of bibliographic services not to be considered as sufficient indexing tools. One current example of Arabic periodicals which offers such services is the Arab Future (المستقبل العربي). This monthly periodical provides a bibliography for the contents of Arabic and foreign periodicals as well as books. The number of Arabic periodicals covered is not mentioned in the bibliography. However, each issue covers not more than 25 periodicals (see issues 118, 151 and 156 respectively published in 1988, 1991 and 1992). A sample of this bibliography is shown in figure 2.8.
قانون وإدارة عامة

دوريات
120 - " массات", مجلة "التعاون الخليجي"، السنة 1991، العدد 2، ص 9-10
121 - "ALARID", مجلة "المستقبل العربي", السنة 2001، العدد 3، ص 11-12
122 - "مجلة"، مجلة "المجلة العربية"، السنة 2002، العدد 4، ص 13-14
123 - "مجلة"، مجلة "المجلة العربية"، السنة 2003، العدد 5، ص 15-16
124 - "مجلة"، مجلة "المجلة العربية"، السنة 2004، العدد 6، ص 17-18
125 - "مجلة"، مجلة "المجلة العربية"، السنة 2005، العدد 7، ص 19-20
126 - "مجلة"، مجلة "المجلة العربية"، السنة 2006، العدد 8، ص 21-22
127 - "مجلة"، مجلة "المجلة العربية"، السنة 2007، العدد 9، ص 23-24
128 - "مجلة"، مجلة "المجلة العربية"، السنة 2008، العدد 10، ص 25-26
129 - "مجلة"، مجلة "المجلة العربية"، السنة 2009، العدد 11، ص 27-28
130 - "مجلة"، مجلة "المجلة العربية"، السنة 2010، العدد 12، ص 29-30

ثقافة

دوريات
131 - "مجلة"، مجلة "المجلة العربية"، السنة 2001، العدد 1، ص 1-2
132 - "مجلة"، مجلة "المجلة العربية"، السنة 2002، العدد 2، ص 3-4
133 - "مجلة"، مجلة "المجلة العربية"، السنة 2003، العدد 3، ص 5-6
134 - "مجلة"، مجلة "المجلة العربية"، السنة 2004، العدد 4، ص 7-8
135 - "مجلة"، مجلة "المجلة العربية"، السنة 2005، العدد 5، ص 9-10
136 - "مجلة"، مجلة "المجلة العربية"، السنة 2006، العدد 6، ص 11-12
137 - "مجلة"، مجلة "المجلة العربية"، السنة 2007، العدد 7، ص 13-14
138 - "مجلة"، مجلة "المجلة العربية"، السنة 2008، العدد 8، ص 15-16
139 - "مجلة"، مجلة "المجلة العربية"، السنة 2009، العدد 9، ص 17-18
140 - "مجلة"، مجلة "المجلة العربية"، السنة 2010، العدد 10، ص 19-20

علوم وتقنية

دوريات
141 - "مجلة"، مجلة "المجلة العربية"، السنة 2001، العدد 1، ص 1-2
142 - "مجلة"، مجلة "المجلة العربية"، السنة 2002، العدد 2، ص 3-4
143 - "مجلة"، مجلة "المجلة العربية"، السنة 2003، العدد 3، ص 5-6
144 - "مجلة"، مجلة "المجلة العربية"، السنة 2004، العدد 4، ص 7-8
145 - "مجلة"، مجلة "المجلة العربية"، السنة 2005، العدد 5، ص 9-10
146 - "مجلة"، مجلة "المجلة العربية"، السنة 2006، العدد 6، ص 11-12
147 - "مجلة"، مجلة "المجلة العربية"، السنة 2007، العدد 7، ص 13-14
148 - "مجلة"، مجلة "المجلة العربية"، السنة 2008، العدد 8، ص 15-16
149 - "مجلة"، مجلة "المجلة العربية"، السنة 2009، العدد 9، ص 17-18
150 - "مجلة"، مجلة "المجلة العربية"، السنة 2010، العدد 10، ص 19-20

Figure 2.8 : sample of the bibliography provided by the Arab Future Journal
2.5 Arabic subject heading lists and thesauri

As revealed earlier, almost all Arabic printed indexes are not generated automatically and human indexers when assigning keywords to articles are relying on Arabic keyword lists which can be either subject heading lists or thesauri. Therefore, the efficiency of such indexes depends to a great extent on the expertise of Arab indexers and on the comprehensiveness and structure of Arabic keyword lists. Since the discussion of Arabic conventional printed indexes requires consideration of the current situation of the available Arabic keyword lists, this section will highlight in two subsections the Arabic subject heading lists and the Arabic thesauri which are currently available for derivation of relevant indexing terms for indexed Arabic articles.

2.5.1 Arabic subject heading lists

In a study introducing and justifying the launch of the latest Arabic subject heading list the Great Arabic Subject Heading List (قائمة موضوعات العربية الكبرى) prepared by Khalifah and al-‘Ayidi (1985a), Arab scholars in the field of Librarianship, Khalifah and al-‘Ayidi (1985b, p.48) said that on the Arab level it is not possible to discuss general Arabic subject heading lists before the year 1978, and that due to their recent appearance most public and academic Arabic libraries, even in well-off countries such as Saudi Arabia, Qatar, United Arab Emirates and Libya, do not have proper subject catalogues for their Arabic collections. As an example, they mentioned the subject catalogues at the University Library of Cairo University, where the availability of Western subject heading lists LCSH (Library of Congress Subject Headings) has made it possible for the library cataloguers to build up subject catalogues to Western Language
Collections whereas the absence of a relevant Arabic subject heading list for university libraries is hindering the building of similar Arabic catalogues for Arabic collections.

In this study, Khalifah and al-‘Ayidi (1985b, p.53) went on to discuss the shortcomings of the existing general Arabic subject heading lists and revealed the irrelevance of most available specialized lists to be considered as standards for indexers to use in indexing the contents of items in Arabic indexes and bibliographies. They finally claimed that the aim of the new Great Arabic Subject Heading List was to introduce a better subject heading list to be used in national, university, specialized and large public libraries all over the Arab world. However, ’Itayyim (1986, p155) mentioned that the authors largely depended on the previous two general Arabic subject heading lists prepared separately by al-Khazindar (1978) and the Cataloguing and Classification Department in Riyadh University (1978), but did not really correct mistakes and defects noted in them. In addition, Ubaidly (1988a) confirmed this fact and also summed up the most obvious problems which directly affect the preparation of "printed indexes for Arabic periodicals" when indexers use the Great Arabic Subject Heading List.

In another study concerned with Arabic subject heading lists and their efficiency to be used in indexing the contents of items other than monographs such as periodical articles, Farsuni (1981) stated that the Arabic Subject Heading List prepared by the Cataloguing and Classification Department in Riyadh University in 1978 did not take into consideration the need of Arabic libraries for specific subject headings to index the contents of their collections of periodical publications. Moreover, Ubaidly (1988b) studied the efficiency of the available three general Arabic subject heading lists and observed the lack of widespread concepts such as التصحر (desertification)
and الذكاء الإصطناعي (artificial intelligence). He also observed that in many cases broad subject headings were used rather than specific ones. For more examples about the deficiency of the examined lists, see the table provided in Arabic with his study.

As far as the structure of the Arabic subject headings is concerned, it is observed that compilers of Arabic subject heading lists and Arab librarians still wallow in constructing natural or reversed subject headings. Librarians like al-Khatib (1988) criticize the Great Arabic Subject Heading List and call for the use of direct unreversed subject headings as it is more suitable in the Arabic natural language. On the other hand, other librarians like Yusuf (1989) find that reversed subject headings are better and more useful. Both groups defend their ideas and support them with relevant examples. Until the proper ideas are resolved, it is relevant to mention that conventional Arabic printed indexes and indexes which derive terms from subject heading and keyword lists in the Arab world are still far behind national and international standards with regard to their frequency and coverage.

2.5.2 Arabic thesauri

The situation of Arabic thesauri is no better than that of Arabic subject heading lists. 'Itayyim (1986, p.156) who surveyed the existing Arabic thesauri, stated that the thesaurus ال-جامعـة is the only Arabic thesaurus which is being used manually and electronically. ال-جامعـة thesaurus, which was published by the Arab League Documentation Centre in 1987 and reprinted in 1993, is a trilingual tool for Arabic, English and French. It was started as a translation of the thesaurus UNBIS of the Dag Hammarskjold Library of the United Nations, but due to the major development which it gained during the preparation it was given this different
Among the retrospective and current Arabic printed indexes, only the two indexes published by Dilmun for Publishing are partially dependent on one of the reported Arabic thesauri. The Islamic Index and Palestine Record partially rely on the thesaurus of al-Jami'a to derive convenient keywords for their subject indexes. The Bibliographical Guide to the Arabic Writings on Libraries and Documentation published by ALECSO might have depended in its latest issue on a specialized thesaurus for Library and Information Science published by the same organization called Thesaurus of the Terminology in Library and Information Science (مكتز مصطلحات علم المكتبات والمعلومات).

In addition to that, there might be other in house indexing services which use general and specialized Arabic thesauri and also there might be some Arabic thesauri which have not been publicized. For instance, Salem (1991) reported two case studies for bilingual Arabic and English thesauri construction in Kuwait. These two case studies were not mentioned by 'Itayyim's survey in 1986 of published and unpublished Arabic thesauri. Salem (1991) described in detail the process of constructing the Arabic Thesaurus in Social, Economic and Political Activities (ATSEPA) which was started in 1982 and the Arab Petroleum Thesaurus (APT) which was started in 1985. According to Salem, progress in both thesauri was disrupted and neither of them was published due to the Iraqi invasion of Kuwait in August 1990. For a current survey of the existing Arabic thesauri, it is advised to see 'Abd al-Rahman (1993), who counted most of them in his paper about the use of "AL-MUSTASHAR" system to establish thesauri in Arabic language.

Unlike the Arabic subject heading lists, Arabic thesauri were prepared by institutions and many of them were supported by governmental bodies and
international committees. Despite its necessity in Arabic, Arabic thesauri are still far behind the national and international standards. Several of them are being used with no indication of regular updating to their descriptors.

2.6 Conclusion

In this chapter, almost all types of Arabic printed indexes which cover the contents of Arabic periodical articles were approached. It is observed that they generally suffer from non-currency and unsatisfactory coverage for periodicals published in the Arabic language. Many of them had started as current indexing tools before they declared themselves retrospective or turned to handling retrospective periodical articles. Reasons behind this may be the dearth of specialized indexers who know Arabic and the conventional indexing processes which are being used.

Although the gap between the publication date of an Arabic journal and its appearance in the available printed indexing tools has been narrowed, the introduction of computer facilities to Arabic periodical indexing has not caused considerable improvement as regard to both currency and coverage. This has been due to the fact that computers are being used only to perform purely clerical tasks such as storing the data and then sorting them out according to prespecified arrangements. The recent Arabic periodical index, the Palestine Record, which depends on computer facilities in formatting its printed output, is in arrears. The issue number 4 of volume 1 of this index which covers periodicals published between October and December 1992 was published in April 1993. Obviously it is in arrears, although it only covered the contents of 342 articles taken from 42 Arabic periodicals.

For Arabic printed indexes to keep current and to be more up-to-date with wider periodical coverage, a better computerized indexing system has to
be adopted. This indexing system should consider the shortage of Arabic specialized indexers and the need to disseminate Arabic literature as cheaply and quickly as possible. Also, it should consider the need to produce cumulated indexes. Therefore, computers should be used to perform both clerical and intellectual tasks. In Arabic indexing, computer facilities have been used only to perform the clerical task: other intellectual tasks that can be carried out by computer packages have never been tried in the process of producing Arabic printed indexes.

Currently indexing systems which allow the generation of printed indexes by means of computer with no or little human intervention are numerous. Several studies about computer techniques which were developed for the purpose of generating printed indexes have been reported in the English literature. Campey's survey (1973) revealed a "total of 118 available index generation packages... covering a wide range of index types".

One sort of index generation package depends on title keywords of periodical articles in order to produce printed indexes automatically. Whether such index generation packages accept Arabic language scripts and whether titles of Arabic articles can be used to produce Arabic printed indexes for Arabic periodicals involves many issues which have to be studied and examined. Therefore, the following chapter is dedicated to discussing the title keyword indexing techniques and the case for Arabic literature.
chapter three

**Keyword indexing techniques and the case for Arabic literature**

This chapter is devoted to revealing the history of title keyword indexing techniques and the various methods used in the generation of title keyword printed indexes in English language. It also casts light on the title word indexing process in the Arab world as well as on the advice that has been offered to non-Arab authors regarding the composition of their article titles in contrast to that available for Arab authors. Then it surveys the functions of Arabic article titles as they are the basic requirement in the application of title keyword indexing techniques on Arabic articles. Finally, it looks into the advantages and drawbacks of using the title keyword indexing techniques for items in English. The question of whether such advantages and drawbacks will be the same for the application of title keyword indexing for Arabic article titles is kept for the following chapters.
3.1 Introduction

Keywords, which are variously termed as significant words, substantive words, catchwords or lead terms can be defined as those words which characterize and reflect the subject matter of a document more than other words in an item. In a proper title keyword index, they constitute the main part of each entry beside the contexts and the reference codes. It is only through them that users can consult the index.

Unlike indexer-assigned subject headings in a conventional index, keywords are derived from the document. The primary aim of keyword indexing is to have access points under each keyword. These may be derived from the entire document, the abstract, the title or any combination of these. This type of indexing technique is not completely new. However, it is within the last three decades that the keyword indexes have become popular and keywords have been considered as useful elements for information handling. Indexers have been accustomed to using lists of subject headings, such as LCSH (the Library of Congress Subject Headings), Sears' List of Subject Headings, and MeSH (the Medical Subject Headings) to index the contents of the items which they store or print in an index for subsequent retrieval and dissemination. They have been depending on predetermined headings accumulated in lists and schedules to catalogue the newly published literature which comes into their keeping and which often includes new technical terms.

An essential difference between keyword indexing, also known as derivative indexing and subject cataloguing, also known as assignment indexing can be seen through their ways of representing the subject matter of the analysed document. For instance, while the principle of keyword indexing
is to derive access points from words in the document, subject cataloguing involves the conceptual analysis of the content of the document and representing it by a set of controlled words or phrases selected by the indexer from the prespecified list of subject headings.

3.2 History of keyword indexing

The aim of this research is restricted to the keyword indexing techniques applied to titles of articles. However, this technique has been applied to the entire document to produce concordances and studies were carried out to investigate its application to abstracts.

3.2.1 Concordances

In fact, the most elementary form of index, where each separate word in a book was listed, is the concordance. Collison (1959, p.17) quoted that John Marbeck made a concordance to the Bible in 1550, and mentioned that Alexander Cruden compiled his concordance in 1737. Taylor (1966, p.32) in his turn cited another concordance to the Bible published in 1577. However, it has been said that Cruden's concordance to the Bible claimed to be the first complete concordance. Collison (1959, p.17) mentioned that Cruden devoted many years to accomplish it, and that his concordance is still in daily use although it is about three centuries old.

It is relevant to state here that concordances have also been compiled to books other than religious. Collison (1959, p.102) cited one concordance to Shakespeare's works. He said "concordances exist for the works of the very great authors. Nevertheless, the works of many writers, either because they are too obscure or because they are too recent, lack any index to their contents. The work of compiling a concordance is laborious and monumental,
but if a good example is used, such as Bartlett's concordance to the works of Shakespeare, it is possible to construct and publish a most useful concordance which will be of great use to oneself and to others interested in the same author.

3.2.2 Abstracts

Caras (1968) says "as a compromise between the title, which does not normally contain sufficient terms, and the full text, which contains too many, abstracts of documents have been considered as a source of index terms". His paper on indexing from abstracts of documents did not report the existence of any printed index being published using keywords from abstracts. He assumed that abstracts contained sufficient terms, and he studied the two kinds of abstract, informative and indicative in order to show how does one select which abstract to use. Previously Slamecka and Zunde (1963) found that 80.4% of the indexer-assigned terms were also contained in the abstracts of the documents taken from NASA's Scientific and Technical Aerospace Reports.

Although Slamecka and Zunde (1963), and Caras (1968) studies have not apparently shown the reflection of their results in printed indexes based on abstract keywords, their usefulness should be obvious for those who consider the online searching of bibliographic databases, such as LISA (Library and Information Science Abstracts) and AIB (Arab Information Bank).

Lancaster (1991, p.227) reported few studies on trials which depended on article abstracts in the production of printed indexes. The first of which was conducted by Trubkin (1979) who described the process of automatically indexing the abstracts in ABI/INFORM (a database in the field of business)

3.2.3 Titles

The history of indexing by each keyword contained in a title can be divided into two main stages. The first began more than one century ago and ended in the late 1940s. The second which has been much more active, started in the early 1950s. The distinction between these two stages has appeared with the emergence of the computers as possible indexing tools.

The first stage, which is believed to have started with the index to WATT's Bibliotheca Britannica in 1824 and Andreas Crestadoro's great work "The art of making catalogues" published in London in 1856, was based on the manual procedures in making such type of indexes. These indexes can be called indexer-produced title word indexes, because they relied only on the indexer's part in both clerical and intellectual efforts.

On the other hand, the second stage indexes have relied upon the computer which has come to supersede the indexer's clerical and intellectual efforts in producing the same type of indexes. Indexes of the second stage are often called machine-produced title word indexes.

In fact the basic principle of indexing each significant word contained in a title was devised long before H. P. Luhn's KeyWord-In-Context principle which is based on mechanizing the production of such type of indexes. An early example of manually manipulated indexes of this type is provided by the index to Watt's Bibliotheca Britannica (Watt, 1824). The basic components which form an index entry in this index are a keyword, which is often taken from the title of the indexed document; the date; the title, with the repeated keyword abbreviated; and the link key (reference code). An example of this index, taken from Craven (1986, p.49) for the title entries "
Peace, Ignominy, and Destruction; a poem" is shown in figure 3.1.

Some years later, a British librarian Andreas Crestadoro (1856) used the rotation of words in titles, claiming that thus the index subject headings would follow the author's own definition of the contents of his book. Crestadoro used a method that he called "concordance of titles" to rotate words in titles for the compilation of a catalogue of the Manchester Public Library. The catalogue consisted of a "list of principal entries" providing full bibliographic details about items; and an index of "subject-matter entries" each consisting of the subject heading, the author's name and the imprint of the item, concluding with the code number referring to the corresponding principal entry of the same item. An example is shown in figure 3.2.

In the introduction to "Manchester Free Library Catalogue" Crestadoro (1864) stated that "an alphabetical condition of titles under author's names does not by itself constitute a catalogue, any more than a collection of men is of necessity an army. The men, when recruited must be organized for warlike operations -- so titles, when gathered together to be useful as a catalogue, must be further arranged for purposes of reference and study". He added that "this is what has been attempted in the 'subject-matter entries'. Everything in the full titles, which was considered likely to be an object of research, has been picked out and collected together in an alphabetical consulting index, where it is hoped, the reader and the student will find all that they may seek".

Another fact about the old use of keywords was stated by Lewis (1964) who mentioned that "German libraries were using the Schlagwort (catchword or keyword) in their cataloguing procedures, one hundred years ago or earlier".
Having explored such an old period of indexing, it can be said that indexing by title keywords was found and practised before the development of the currently used title keyword indexing techniques made popular by H. P. Luhn's KWIC (KeyWord-In-Context) system, and that the indexing process with the use of context to clarify the keywords' meaning was in practice in the old European libraries.

The second stage of the history of title keyword indexing began in the early 1950s when many specialists started to look at the computer as a possible indexing tool. The first system to be used for machine manipulation and compilation of rotated title-and-term-index records has been in productive operation since 1952. Mary Veilleux had been operating a rotated title-words indexing system at CIA (the Central Intelligence Agency) since 1952, and first reported this fact in 1961 at the American University's third Institute on Information Storage and Retrieval. Veilleux (1961) said "Consumer demand balanced against availability of manpower and machine time were the factors which led to the establishment of the permutation title word indexing project in 1952 ". Afterwards, Stevens (1965, p.41) mentioned that in 1958, Hans Peter Luhn and Joan Citron and his associates Lewis Hart and Herbert Ohlman independently developed a rotation index to the preprints of the International Conference of Scientific Information with the help of a machine, and distributed it at the conference at Washington in that year. Also in the same year, Baxendale (1958) published the results of her experiments in automatic indexing involving scanning of topic sentences, syntactical deletion processes and automatic phrase selection.

It is worth mentioning here, that Luhn (1960) developed a rotation index on his own and called it KWIC (KeyWord-In-Context), and his work which was distributed during the International Conference was based on titles.
Whereas, Citron and his associates (1958) developed a rotated keyword index based on titles, subtitles, section and table headings, figure captions and selected sentences, or phrases taken directly from the text.

After this glance at the history of indexing by each keyword contained in a title, it can be said that although the identity of the developer may be in doubt, no one can dispute the fact that the popularizer and the person who deserves the credit for making KeyWord-In-Context a common term and a popular form of index was Hans Peter Luhn. Actually, the significance of Luhn's KWIC index is based on the fact that it is produced automatically by machine, affording speedy compilation, accuracy and completeness.

1. DESTRUCTION. - 1796. Peace, Ignominy, and D.; a Poem. 546 o
2. IGNOMINY. - 1796. Peace, I., and Destruction; a Poem. 546 o
3. PEACE. - 1796. P., Ignominy, and Destruction; a Poem. 546 o

Cross-references and notes are also included; for example,
PEACE, rest, quiet

Figure 3.1 : sample of Watt's Bibliothica Britannica, taken from Craven (1986)
C A T A L O G U E
OF THE
MANCHESTER FREE LIBRARY.
REFERENCE DEPARTMENT.

SUBJECT-MATTER ENTRIES OR CLASSIFICATION.

The numbers at the end of the lines refer to the principal entry of each work. The small numbers at the beginning stand as many words in the preceding line to be repeated.
3.3 Computer characteristics

In fact, the preparation of both types of indexes, assignment and derivative, was a tedious and time-consuming process until the emergence of the computer which gave indexing practice a new approach regarding the facilities which it offers and the characteristics which it has in this respect. The computer as an indexing tool has been a great assistant in the production of printed indexes and online searching for information.

Before considering the role of the computer in the indexing process it is useful to clarify the incentives that encouraged the use of the computer and its characteristics.

The huge quantity of information that has to be treated constitutes one of the basic motives why specialists have had to evolve new methods and techniques of information handling. Other motives are that the tempo of life is generally faster and techniques which were sufficient and adequate in the past are no longer so. Black (1962) says that "conventional methods for indexing have within recent years begun to show signs of strain, and in some cases breakdown, due to the tremendous increase in the volume and complexity of technical literature. Advances in chemistry, atomic physics, electronics and aeronautics have distorted and blurred the existing systems of classification. The logicality of hierarchical breakdown is often non-existent. The overlapping of technologies into many disciplines imposes a great strain on existing systems". On the other hand, the computer itself offers a means of processing huge quantities of information at very high speed.

After stating these incentives, it is therefore obvious that specialists have had to consider the expediency of computers in executing the basic processes of information retrieval systems. These processes which differ from
each other are often called intellectual and clerical operations. For instance, deciding which entry points an item should be indexed under is a completely intellectual operation, whereas the actual mechanics of putting an entry in a file are not. Intellectual and clerical operations and the distinction between them become much more important when the use of computer-based systems is considered. Besides the reasons cited by Black (1962) why conventional methods are showing signs of strain, more can be added with respect to the characteristics of computers. Four are mentioned below.

3.3.1 Speed

Generally, computers can store and retrieve information at a very high speed. Many tasks that would take an indexer days to accomplish, a computer can complete in a matter of seconds. A large modern computer can perform millions of arithmetic operations per hour.

3.3.2 Accuracy

This reason can be considered as a major factor in using computers for the purpose of indexing. Actually, there is a clear distinction between a job which has been done by an indexer and a job which has been done by computer. Usually, indexer's efforts are subject to careless or "human" error, whereas a computer can process data with uncompromising accuracy. Studies of the consistency of human indexers who index the same document (inter-indexer) is rather low, and even the agreement within the same indexer who indexes the same document at different times (intra-indexer) is low as well. Zunde (1965), for example, reports inter-indexer consistency coefficients ranging from 0.158 for six indexers to 0.453 for two indexers, and an intra-indexer consistency coefficient of 0.661.
Furthermore, it is generally accepted that computers do not make mistakes because they merely follow instructions. Erroneous computer outputs or outcomes are invariably due to faulty logic on the human side.

3.3.3 Reliability

While the quality of an indexer's work may suffer or be affected as a result of distraction, fatigue or illness, the computer does not get bored and is extremely reliable.

3.3.4 Cost

Previously, the costs of having or employing computers have been such as to prevent their use in information retrieval. But, the new developments which have been done in the computer industry and technology in this decade have reduced costs to the level where they are competitive with clerical or manual effort. Foskett (1982, p.38) saw differently when he said "the advantage in turning to computers does not lie in reduction of costs, but in the fact that for about the same cost we can get a great deal more output; in other words, we spend no less but we get more for our money". Actually, computer processing can give us a variety of output products from the same input.

3.4 Title-derivative indexes generated by computers

The entry of the computer to the field of information processing has brought about several changes in the world of indexes and their kinds. Title keyword indexing techniques using the facilities provided by computers have led to the generation of many types of index but basically they can be divided into three:
3.4.1 Indexes using the rotation of title KeyWords-In-and-Out-of-Context

All the computer generated printed indexes which use the rotation technique in their production go under this category of indexes. This type of indexing technique involves computer rotation of title keywords and their subsequent alphabetical ordering. The aim of this computer-based system technique is to shift each significant word in a title to a fixed indexing position, in order finally to get a listing of each such word together with its accompanying context. The title keyword rotation technique is simple and straightforward. The entire indexing process is performed automatically once the data have been recorded in the computer-based system. Entries of bibliographical information such as title, author, source and date are keyboarded, so that these data may be processed by the computer program to produce the final index.

After being read by the computer program, the entered data are processed against a "stopwords list" in order to eliminate from further processing the insignificant words. Afterwards, all the remaining presumably significant words, which have not been included in the "stopwords list" in the computer memory, are then taken one at a time to an indexing position that can be called the indexing window as well, where they are sorted into alphabetical order. The position of the indexing windows varies from one indexing type to another, and the space allocated for the index entries is also varied accordingly. The most known and reported types of indexes which use the rotation technique are KWIC, KWOC, WADEX, AKWIC, and OKWIC. These types are described separately in the following sections.
3.4.1.A KWIC (KeyWord-In-Context)

In the KWIC type indexing, the outcome of using the rotation technique is a listing of title keywords sorted alphabetically. In the process of producing a KWIC type printed index, each keyword in an article title becomes an access point, but instead of the conventional appearance of an entry point at the left hand side of the index page, the access point (keyword) appears in the middle with its surrounding context, in the permitted space in a given format.

The indexing window in a KWIC type printed index is usually located at or near the centre of the line entry with several extra spaces to the immediate left. Examples of the generation process and the KWIC index are respectively shown in figures 3.3 and 3.4.

3.4.1.B KWOC (KeyWord-Out-of-Context)

The KWOC indexing technique is based on extracting a keyword from its context and following it by the complete title in its normal order. Its window, where the keywords are displayed, is either at the left most margin, as shown in figure 3.5, or at the left margin immediately above the index line entry, as shown in figure 3.6. Each title in a KWOC index appears in the final list in a number of places each according to one of the significant words it contains. Titles which have only one significant word will appear once.

3.4.1.C WADEX (Word and Author inDEX)

WADEX is an index relying on a KWOC format consisting of keywords which appear either in the title of a document or in its author's names. In this indexing technique the names of authors are included with
titles and considered as significant words. Each significant word, which is alphabetized in a column left of the title, out of context, is accompanied by a complete title, author's name and a reference code. An example of WADEX index is shown in figure 3.7.

3.4.1.D AKWIC (Author and KeyWord-In-Context)

AKWIC, as has been defined by Juhasz et al. (1969), is an amplified KWIC index with the authors' names of the documents added to their titles. The surname of the author is added first, then it is followed by his initials which have no space or blank between them.

AKWIC and WADEX have the similar advantages of using the authors' names as index terms. However, in this respect, AKWIC has a drawback which WADEX has not. It is apparent that as a result of adopting the KWIC (KeyWord-In-Context) technique, the AKWIC index line entry is bound to be truncated each time it exceeds the limited space allotted. So that the author's name may be split and loses its value in the context. A sample of this index is shown in figure 3.8.

3.4.1.E OKWIC (Online KeyWord-In-Context)

In the Online KeyWord-In-Context index, the user selects the articles to be indexed through a regular online search. Cuadra (1978) stated that "the user can cover any number of databases, and the Online KWIC program will merge them into a single unified KWIC index to items printed offline in a separate bibliography". According to Cuadra, the user's reaction towards the efficiency of Online KWIC was excellent. An example is displayed in figure 3.9.
It is observed that OKWIC, as Bachir (1987, p.94) refers to it, does not suppress the insignificant words. The letter " A " appears in the indexing window and sequenced alphabetically with other keywords.

Figure 3.3: the generation process of a KWIC index
Figure 3.5: sample of a KWOC type index with indexing window at the left margin
COLUMBUS (CONTINUATION)
AN ANALYSIS OF PRINCIPAL APPRAISALS OF THE TEACHING SERVICE OF LAY TEACHERS IN THE ELEMENTARY SCHOOLS OF THE DIocese OF COLUMBUS. DEBES-EB-APA

COMEDIES
A COMPARISON OF TWO OF THE SERIOUS PLAYS OF CHRISTOPHER FRY WITH TWO OF HIS COMEDIES ON THE BASIS OF CERTAIN CRITICAL VIEWS AND EVALUATIONS. WALKER-EB-CTS

AN INVESTIGATION OF THE TEN LARGEST GOSPEL INCOME MUSICAL COMEDIES PERFORMED IN NEW YORK CITY BETWEEN 1948 AND 1950 AND THEIR OUTLET FOR ADVERTISING EXPENSES. KNAPP-EB-11L

COMEDY
A COMPARATIVE ANALYSIS OF THE COMIC ELEMENTS AS FOUND IN A SHAKESPEAREAN COMEDY, AS YOU LIKE IT AND TRAGEDY, HAMLET. BARKER-EB-CAH

THREE ROBINS: A MUSICAL COMEDY BOOK IN TWO ACTS (INCLUDING SATISFICATIONS FOR SYRETY). HARRA-EB-YLR

GOVERNOR, MY GOVERNOR, A COMEDY IN THREE ACTS. CATCH-EB-SNG

SHEET RELIEF, A MUSICAL COMEDY. LACI-EB-SAR

AFTER FIVE, AN ORIGINAL COMEDY. NODA-EB-APO

OXY, AN ORIGINAL MUSICAL COMEDY. SARBU-EB-DOR

HISTORICAL DEVELOPMENT OF MUSICAL COMEDY IN AMERICA. TORADA-EB-DEN

AN EDITION OF LORENZO DI FILIPPO STROZZI COMEDY, LA VIOLANTE, WITH AN INTRODUCTION TO THE COMEDIES ERRATA OF THE CINQUECENTO. TIMOT-EB-9L0

COMFORT
FACTORs IN EVENING CARE IDENTIFIED AS CONTRIBUTING TO PATIENT COMFORT. COLLAM-EB-PEC

COMEDY
A COMEDY IN THE COMIC FORM. SOUSA-EB-9UCA

THE NATURE OF THE COMIC ACTION. SURIJU-EB-9CA

AN ANALYSIS OF THE COMIC ELEMENTS IN TWELFTH NIGHT. HARDMA-EB-ACE

A COMPARATIVE ANALYSIS OF THE COMIC ELEMENTS AS FOUND IN A SHAKESPEAREAN COMEDY, AS YOU LIKE IT AND TRAGEDY, HAMLET. BARKER-EB-CAG

THE STYLISTIC MEANING OF EPIPHRONIC EUPHOMIA IN LES COMEDIES AGILELS. NOIMA-EB-9SHE

BARDUB
A COMPARISON OF TWO NOVELS, SEARRDANS ROMAN CORISQUE AND GAUTIERS (EPITHESE FRACLESSE. TREAG-EB-CTH

COMMAND
PERMISIONS BY APDO GRADUATE ENGINEERS OF THEIR UTILIZATION IN THE AIR FORCE SYSTEM COMMAND. LLPPC-EB-PAC

THE CONSTRUCTION OF A TOOL TO ASSESS THE ATTITUDES OF ROMAN CATHOLIC NURSES TOWARD THE MANAGERIAL PRINCIPLES, UNITY OF COMMAND. CASMA-EB-ZCTA

COMMANDERS
A STUDY OF THE READINESS OF TROOP COMMANDERS TO USE THE ARMY RENTAL HYGIENE CONSULTATION SERVICES. HAILED-EB-9RT

COMMERCE
THE DIVINE COMEDIES IN FRENCH- A STYLISTIC STUDY OF LONGHON TRANSLATION. DODOA-EB-4DCP

AN EDITION OF LORENZO DI FILIPPO STROZZI COMEDY, LA VIOLANTE, WITH AN INTRODUCTION TO THE COMEDIES ERRATA OF THE CINQUECENTO. MELT-EB-9L0

COMMERCE
THE COMMERCE CLAUSE, EVOLUTION IN ITS INTERPRETATION SINCE 1901. PERL-EB-9CE

RAILROAD DECISIONS OF THE INTERSTATE COMMERCE COMMISSION, THEIR GUIDING PRINCIPLES. BISRMA-EB-901

COMMERCIAL
A COMMERCIAL CENTER FOR ALEXANDRIA, EGYPT, U.A.R. MAASS-EB-9CA2

A CATALOG OF AMERICAN FOLK-MUSIC ON COMMERCIAL RECORDINGS AT THE LIBRARY OF CONGRESS, 1925-1940. SPOTT-EB-9CA

COMMISSION
THE INFLUENCE OF THE PRESIDENT'S EDUCATIONAL CONFERENCES AND COMMITTEES ON THE FEDERAL GOVERNMENTS PARTICIPATION IN AMERICAN EDUCATION. MFAPE-EB-9IPE

COMMITMENT
THE ASSOCIATION OF OCCUPATIONAL DISORDER WITH DIAGNOSIS, PREVIOUS TREATMENT, TYPE OF COMMITMENT AND DISPOSITION OF MENTAL PATIENTS. CINDY-EB-4A0

FAITH AS COMMITMENT IN THE LIGHT OF RECENT THEOLOGY AS SEEN IN CATECHISMS PROGRESSIVE. HOPIT-EB-4TC

A STUDY IN CHRISTIAN COMMITMENT TO THE MODERN WORLD, THROUGH AN EXPERIENCE OF MISSIONS AND NEEDS. GIBOR-EB-9SCE

TEACHING THE APOSTOLIC COMMITMENT IN THE CATECHUMENATE. BARBA-EB-9FC

Figure 3.6: Sample of a KWOC type index with indexing window above the index line entry

106
Figure 3.8: sample of an AKWIC type index
3.4.2 Indexes using the rotation of enriched title

KeyWords-In-and-Out-of-Context

In an attempt to improve retrieval and to express the subject matter of an item more precisely with the use of rotation technique, additional significant words are either derived from the text of an article or from its abstract and added by the help of a specialist indexer to the title of the article. The editing process of the title, if it is required, is done at the time of title enrichment while filling the input work-sheet. Two types KWWC and ETKW of the derived printed indexes which are reported in the literature are cited below.

3.4.2.A KWWC (KeyWord-With-Context)

KWWC is a variation of KWIC within which the full context is not repeated with each entry point. Only the segment which is pertinent to the entry point is provided. Ghosh (1981, p.187) mentioned that in order to denote precisely the subject, in some cases, a secondary aspect is used with the primary one, separated by a punctuation mark. An example of this index is shown in figure (3.10).

3.4.2.B ETKW (Enriched Title-based-KeyWord)

The Enriched Title-based KeyWord (ETKW), which is sometimes referred to as KeyWord-Augmented-in-Context (KWAC) is, according to Rajendran (1986), used by about 500 members of ERIC (Educational Resources Information Center) located in more than 95 countries.

The way in which an original title is enriched and edited is shown in figure 3.11. The words which appear after the # sign are additional significant words resulting from the enrichment process which is done by indexer
intervention in order to enhance the relevance of this article title.

In ETKW each pertinent word in the enriched title is indexed including the additional keywords. A dash between two significant words means that these words will appear in the proper index as a single term. After suppressing insignificant or stopwords, the title given would be indexed under the following entry points:

**Fuel-saving ; Cooking-stoves ; Pakistan ; Design ; and Woodstoves.**

A sample of the final ETKW output format taken from Rajendran (1986) is shown in figure 3.12.

The KWWC and ETKW are not the only derived printed indexes which enrich inadequate titles. B.A.S.I.C. printed index (Biological Abstracts Subjects In Context) which is also derived from the title words supplied by authors, is supplemented by significant terms provided by the editorial staff at BIOSIS (Bio Sciences Information Service). Like ETKW, Zabriskie and Farren (1968) reported that the supplemental index terms are selected from the body of the abstract and from the text of the original item.
Title: Validity of diffusion theory in radio frequency breakdown in molecular gases in longitudinal magnetic field 300

KWWC entries:

DIFFUSION THEORY
  Validity in radio frequency breakdown in molecular gases 300

MOLECULAR GAS
  Validity of diffusion theory in radio frequency breakdown 300

RADIO FREQUENCY BREAKDOWN
  in molecular gases 300

Figure 3.10: Sample of a KWWC type index
Original title of a new article:  
A new fuel saving cooking stove from Pakistan

Enriched title of the article:  
A new fuel saving cooking stove from Pakistan # design woodstoves

Enriched title of the article edited with underlined character:  
A new fuel-saving cooking stove from Pakistan # design woodstoves

Figure 3.11: sample of an ETKW type index showing the process of title enrichment

Figure 3.12: sample of the final output of a ETKW type index, taken from Rajendran (1986)
3.4.3 Indexes employing title keywords coordination

Rather than using the rotation technique, indexes of this category employ the permutation technique to generate their title keyword alphabetical listing. The distinction between the rotation and permutation techniques is clarified in section 1.6 of this thesis. The following sections are devoted to such types of printed indexes which are most reported in librarianship and information studies literature. These indexes are the PSI, DOKWIC, and PANDEX.

3.4.3.A PSI (Permuted Subject Index)

PSI is a permuted title-word index to the article titles processed for the SCI (Science Citation Index). It was designed in 1964 at the ISI (Institute for Scientific Information). The reason behind the creation of PSI was the need for a solution to the lack of a starting search point. This problem is commonly faced by users of the CI (Citation Index) section of the SCI (Science Citation Index) because of their limited knowledge of authors' names which constitute a starting point to their search.

PSI has come to provide a path not for the typical scientist-users of the Citation Index section, but for those who have no keys to enter it. In the PSI, which is a companion to the SCI, each significant word in the title is permuted by a computer program in order to produce all possible pairs of terms. For a title containing "n" significant words, there will be n(n-1) possible word pairs created by permutation. Garfield (1976) says that after suppressing the insignificant words, this usually generates about 42 word-pairs for the typical seven word title. Every significant word in the title of an article indexed in PSI is potentially both a primary term which is a main
entry point and a secondary term which is called by ISI a co-term. An example of PSI is displayed in figure 3.13.

3.4.3.B DOKWIC (DOnuble-KWIC coordinate index)

The Double-KWIC coordinate indexing technique is a variant of the KWIC indexing technique. It has been adopted by Petrarca and Lay (1969).

The DOKWIC index is constructed by extracting the first significant word in a title as a main index term and replacing it by an asterisk in order to locate its position in the context. The next step is to cycle the remaining significant words in the title to be displayed as a subordinate term under the main index term, using a wrap-around technique. Afterwards all index entries are sorted alphabetically by main index entry, and within main entry by subordinate significant words. An example is shown in figure 3.14.

3.4.3.C PANDEX index

PANDEX is a development of the KWOC index. Feinberg (1973, p.144) says it relies on selective coordination of title words. The computer program provides for the selection of pairs of words by means of a grammatical algorithm.

The chosen word-pairs are emphasized by boldface in the context whereas entry points are printed in capital letters in boldface as well. Figure 3.15 shows an example of it.
Figure 3.13: sample of a PSI type index
NOMENCLATURE

BIOCHEMICAL * = ..................THE ORGANIZATION AND FUNCTIONING OF 72
CARBOHYDRATE * = ................................................................. 78
CHEMISTRY * = ............................................................................ 64
FLUORINATED MOLECULES * = ..................................................THE * OF HIGHLY 82
FUNCTIONING OF BIOCHEMICAL * = ..................................THE ORGANIZATION AND 72
HIGHLY FLUORINATED MOLECULES * = ..................................THE * OF 82
INORGANIC * IN 1966: PROGRESS AND PROBLEMS = ............. 67
MOLECULES * = ...............................................................................THE * OF HIGHLY FLUORINATED 82
ORGANIC CHEMISTRY * = .............................................................THE * OF 64
ORGANIZATION AND FUNCTIONING OF BIOCHEMICAL * = ...........THE 72
POLYMER * = ..............................................................................SOME PROBLEMS IN 74
PROBLEMS * = ................................................................. INORGANIC * IN 1966: PROGRESS AND 67
PROBLEMS IN POLYMER * = ..........................................................SOME 74
PROGRESS AND PROBLEMS * = ..........................................................INORGANIC * IN 1966: 67

Figure 3.14: sample of a DOKWIC type index
3.5 Title keyword indexing and Arabic articles

Having gone through the most known and reported English title keyword printed indexes which are formatted and generated by means of computer, it is relevant to devote some sections of this chapter to revealing the relationship between the title keyword indexing techniques and the articles of Arabic periodicals. Therefore some of the following sections will reveal and discuss the functions of the titles of Arabic periodical articles and their actual usage and the use of their keywords in the Arabic indexing products.

3.6 The Arabic title keyword indexing tools

Apart from the recent appearance of the Contents of Arabic Periodicals (1989) (محتويات الدوريات العربية), the issue of using title word indexing techniques to produce printed indexing tools does not seem to have been raised before. CAP (the Contents of Arabic Periodical) has been the only indexing tool which depends on titles of Arabic articles. However, its introduction to Arabic literate users as an indexing product came before any feasibility study was conducted on the adequacy of Arabic titles for such use.

As for the issue of producing an online title keyword index in the Arabic language, al-Surayyi (1993, p.322) mentioned in a paper introducing the use of "IBN ALNADEEM" computer integrated system at the Institute of Public Administration Libraries, that the titles file of the items existing in the Institute's libraries was designed in accordance with the permutation technique. Al-Surayyi (1993, p.333) provided an example showing the way in which titles are arranged when search results are displayed on a computer screen. Figure 3.16 shows an example taken from al-Surayyi's paper.
It is believed that the Arabic literature has not reported any other such Arabic indexing tools. In line with that, the manageress of ALDOC (the Arab League Documentation Centre) Fari'ah al-Zahawi (1990) confirmed in a written communication that, as far as she knows, this type of indexing has not yet been used in the Arab World because of the characteristics of the Arabic language.

Figure 3.16: sample of IBN ALNADEEM search result
3.7 Instructions to authors about titles

In this section, some attention is given to the way in which Western and Arab authors deal with titles of their articles. Attention is also given to whether authors who publish in Arabic and English journals are supplied with instructions on how to compose relevant titles and whether they are aware of their potential use in indexing and information retrieval.

3.7.1 English language journals and books

The first specific instructions regarding the composition of adequate titles to be used for title keyword indexing techniques were given to authors who presented scientific papers to the 26th annual meeting of the American Documentation Institute held in Chicago in October 1963. Herner (1963, p.102) advised them that "the title must contain at least six significant words", and predicted that "as authors become aware of indexing, indexing systems, and how their publications are likely to be indexed and searched, they are bound to react consciously or unconsciously in making their writing conform to retrieval vocabularies". At the same conference Kennedy (1963, p.134) urged authors to consider the title as a one sentence abstract and to avoid uninformative words. The proceedings of this conference show that many of the papers presented reflected interest in this approach.

With the growth of literature and corresponding increase in keyword indexing, the importance of titles has increased and because of this increased importance, advice to authors to try and put as many significant and representative keywords as possible into the titles to their articles has continued to be offered.

The Royal Society in England (1974, p.5) has published "general notes
"on the preparation of scientific papers in which the following quotation was found "the title should be specific and as brief as is possible, consistent with giving information, by the use of keywords, that can be useful in indexing and information retrieval ". Afterwards, Mullins (1977); O'Connor and Woodford (1977); Huth (1982); Day (1983); and Hamp-Lyons and Courter (1984) wrote a great deal on how authors should construct the titles of their articles.

Mullins (1977, p.173) advises authors to write accurate titles and to avoid titles that lack accurate descriptors. O'Connor and Woodford (1977, p.47) ask the authors to try to choose the words they would think of if they were looking for their articles in a subject index and if possible to begin their titles with a significant word, cutting out the "meaningless phrases" (stopwords) such as "observations on...", "a study of...", and similar expressions ". Huth (1982, p.76) says "in preparing the title, the careful author makes sure that the title carries the irreducible number of terms needed to accurately describe the content of the paper ". Day (1983, p.13) stresses that "it is fundamentally important that the author provides the right keys to the paper when labelling it. That is, the terms in the title should be limited to those words that highlight the significant content of the paper in terms that are both understandable and retrievable ". Hamp-Lyons and Courter (1984, p.93) state three major areas which authors must consider in writing titles for research papers. These areas are "the form of the title, the accuracy of word choices and grammar ". They add that titles should be as specific as the focused topic chosen by the author.

In addition to this advice, some English journals in the life sciences as well as in the humanities give instructions to their authors in writing titles for their articles. In most journals these instructions are found in sections
variously named "instructions to authors", "instructions to contributors", or "list of instructions".

According to Buxton (1979, p.14) the most detailed early set of instructions was found in the *Journal of Biological Chemistry*: "a clear and informative title is very important in bringing a paper to the attention of readers. The title should be as short as is consistent with clarity; in most instances two printed lines are adequate. The title should not include chemical formulas or arbitrary abbreviations, but chemical symbols may be used to indicate the structures of isotopically labelled compounds. Bear in mind the increasing use of titles in the construction of certain types of indexes, e.g. *Chemical Titles, Biological Abstracts...*. Other samples of such instructions taken from Feinberg (1973) are displayed in figure 3.17.

Such mentioned advice and instructions were given to Western authors to draw their attention to the importance of constructing informative and representative titles, so that their articles could be retrieved and located by users who consult title keyword indexing services. It is very clear from recent English studies that these advice and instructions have influenced the writing style of Western authors who subsequently started to provide more adequate titles. Chapter four of this thesis will cite studies which show that English article titles have become more informative and their numbers of substantive words have increased since the introduction of keyword indexing techniques.
<table>
<thead>
<tr>
<th>Source</th>
<th>General Instructions</th>
<th>Specificity</th>
<th>Brevity</th>
<th>Context</th>
<th>Avoid</th>
<th>Amplification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kennedy</td>
<td>Consider title as one-sentence abstract. Use words rather than symbols.</td>
<td>Select terms as specific as context and emphasis of paper allows.</td>
<td>Balance against descriptive accuracy and completeness.</td>
<td>Sufficient to show relationships.</td>
<td>Uninformative words.</td>
<td>Assign filing subjects for additional subject access.</td>
</tr>
<tr>
<td>Mitchell</td>
<td>Professional and technical articles must indicate content.</td>
<td>Be as specific as the article.</td>
<td>Brevity and succinctness desirable.</td>
<td>Cuteness, wit and strainning for effect.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tetrahedron</td>
<td>Keywords should indicate some special quality or content of communication.</td>
<td>Do not use words with very broad significance.</td>
<td></td>
<td></td>
<td></td>
<td>Amplify insufficient title with 6-8 words from recommended categories.</td>
</tr>
<tr>
<td>Style Manual for Biological Journals</td>
<td>Clarity and conciseness essential.</td>
<td>Make title specific and informative.</td>
<td>Make title short.</td>
<td>Abbreviations, unnecessary &quot;the's&quot; and openings.</td>
<td>Where appropriate, nature of study, experimental organism, technical approach.</td>
<td></td>
</tr>
<tr>
<td>Brandenberg</td>
<td>Balance computer requirements against human scanning habits. Man-machine methods may present conflicts regarding significant words and forms of words.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Efficiency of titles should be assured by author and editor.</td>
</tr>
</tbody>
</table>
3.7.2 Arabic language journals and books

With regard to Arab authors, it is believed that, apart from those who were educated in the West and had the opportunity to read and to publish in Western periodicals in which instruction to authors is supplied, none has been really aware of the importance of titles in the retrieval process.

The only course book to which this thesis could refer on the importance of titles in Arabic literature, is written by a former journalist and lecturer at the Lebanese University. al-Bustani (1982) wrote that course book for a special course to the students of journalism. This course book, although it reveals the importance of titles of newspaper stories, is completely irrelevant to scholars who publish in research and academic journals. On the contrary, it offers suggestions on how to write attractive evocative titles to stories in popular journals. Other books which deal with writing scientific research papers and doctoral theses discuss many relevant aspects of such subjects but never how titles should be constructed and composed. Badr (1979, p.186) only mentioned that titles should not be vague. He stated that when talking about the structure of the title page of a research paper. 'Alam al-Din (n.d.) and Shalabi (1992) wrote two extensive books on how to write doctoral theses and papers without considering the importance of titles and their potential use in information retrieval.

As far as Arabic academic journals are concerned, they often do not provide published lists of instructions to their authors, and if there are any, these lists are restricted to general information about the publishing institution, the aims of the periodical and the method of correspondence with the editor. However, some periodicals have recently started to provide an almost full list of instructions and guidance to their authors on the
organization of the article, the length of its abstract, and the references. Some of these journals are: Studies (دراسات), Arab Gulf Journal of Scientific Research (البحوث العلمية الخليجية العربية), Jordan Medical Journal (الجلة الطبية الأردنية), and Saudi Medical Journal (الجلة الطبية السعودية). Unfortunately, among those periodicals which have started to pay attention to the bibliographic information content of their articles, only one gives instructions to its authors regarding the construction of titles. In order to publish an article, the Arabic Journal of Administration (الإدارة) requests a brief but accurate and informative title. The third condition in this journal's publishing criteria states that "title should be brief and accurate written in correct Arabic and in a clear style using understandable expressions ".

So, it seems that titles of Arabic journal articles are apparently assigned in the ignorance of their potential usefulness in bibliographical control and the information transfer process. However, the situation is bound to change should title-derivative printed indexes for Arabic articles prove to be feasible because then as a basic requirement for title keyword printed indexes to be produced by computer the Arabic titles will become a focus of interest. The Arab indexers would be assisted by Arab authors who will no doubt then be playing a crucial role in determining the quality of their titles and ultimately the quality of the derived indexes. Moreover, the editors of the Arabic journals will also contribute by setting up guidelines for authors in the composition of their titles.

3.8 Functions of titles of Arabic periodical articles

If the application of title keyword indexing techniques enhances the importance of titles and increases the number of indexing tools, its absence
does not mean that they have no usefulness. This section reveals the places in
which Arabic titles appear and are being used for document and subject
retrieval purposes, as well as indicating the subject matter of their
Corresponding articles.

3.8.1 In the primary journals

3.8.1.A In content lists

The contents lists of individual issues of most Arabic periodicals appear
prominently on the contents page, and show articles' titles, authors' names
and page numbers. However, some journals reprint the contents lists on the
cover without mentioning the article's page number and sometimes the
author's names. In Arabic periodicals, usually titles on covers are incomplete.

No studies have discussed the use of contents lists in Arabic, but
Gushee (1968) mentioned that chemists go to the table of contents of English
periodicals to find the articles they want to read and then go to those articles
and read them. Kuney and Weisgerber (1970) found in their study on
utilization of the Journal of Organic Chemistry that 76% of its readers read its
table of contents on receipt and that 23% also look at the author index.

Apart from their use in the initial scanning of periodicals, tables of
contents are very useful for specific article retrieval when readers can
remember the name of the journal in which the required article was
published. Examples of Arabic periodicals which reprint the articles' titles on
their covers are: The Arabic Journal of Information Science (المجلة العربية
للعلومات), Figure 3.18; World of Thought (عالم الفكر); and Dialogue
Forum (منبر الحوار).
Figure 3.18: Cover of the Arabic Journal of Information Science showing titles of articles in the issue
3.8.1.B In reference lists

Rzuwqi (1985, p.79) claimed that quite a large number of Arabic periodical articles do not have reference lists, and if they do then the bibliographic information in these lists is incomplete. Rzuwqi's claim is not always accurate, currently reference lists in Arabic journals are being given more attention. Information on how to organize and what to include in them is found in every journal that provides a list of instruction to its authors. For instance, the *Jordan Medical Journal* (الجلة الطبية الأردنية) instructs its authors that the reference list should include only the cited publications in the text and that "cited references should include the author's surname with initials, full title of article, journal abbreviation, volume number, first page, and year". Moreover, some periodicals which omit the lists of instructions consider the full or incompleteness of reference lists as a basic requirement for accepting or refusing an article. The *Arabic Journal of Information Science* (الجلة العربية للمعلومات) has most of its articles terminated with reference lists, and the *Arab Future* (المستقبل العربي) has chosen to embody cited references in the footnotes sections.

The importance of reference lists has been considered and studied in the West long before they first appeared in Arabic articles. Garfield (1963) believed that references in journals could avoid many unnecessary interlibrary loans; Wood and Bower (1970); Hakulinen (1974); and van Styvendaele (1977) undertook studies which showed that most of the requests for articles arise from the reference lists of journal articles.

3.8.2 In the secondary services
3.8.2.A In the reproduction of contents pages
3.8.2.A.1 Contents of Arabic Periodicals (CAP)

Garfield (1980, p.686) stated that the reproduction of contents pages of journals has been with us since 1958. Yet CAP (Contents of Arabic periodicals) is the first equivalent to contain photocopy and facsimile contents pages from primary Arab journals with a title keyword index. However, unlike CC (Current Contents), CAP does not contain authors' addresses and no study has been undertaken yet to measure its capacity as a source of reprint requests. Although it was first published in late 1989, CAP is considered as the first Arabic equivalent to CC. With regard to other such Arabic services, Bachir (1991) mentioned in reviewing CAP that the reproduction of contents pages of Arabic journals had been in the Arab world before 1989. The Hebrew current awareness publication Current Contents of periodicals on the Middle East has been published by Tel Aviv University since 1980. It includes photocopies of Arabic periodicals' contents pages alongside other language publications. ALDOC (the Arab League Documentation Centre) has been producing a trilingual Periodicals Contents (محتويات الدوريات) in Arabic, English and French since 1986. Another similar service which was not mentioned in that article is Guide and Contents of Libyan Periodicals (دليل الدوريات الليبية ومحتوياتها). This bibliographical tool was first published in 1980 by the Libyan Jihad Centre for Historical Studies. Unlike CAP, none of the above mentioned bibliographical services provides title keyword indexes and they merely include photocopies of contents pages. Therefore, CAP is considered the first of its kind among other such Arabic services in terms of filling a void and disseminating Arabic literature. However, Auchterlonie (1990b) stated that "as a current-awareness tool it fulfills a useful function, although it does need
a broader spectrum of literary and academic periodicals ".

3.8.2.A.2 In journals and daily newspapers

Arabic primary journals such as World of Books (عالم الكتب) of Saudi Arabia specializing in publishing and librarianship, and primary general newspapers such as al-Hayat (الحياة), sometimes contain reprinted contents pages of other general and specialized journals, but without mentioning the page numbers of the articles provided in the original journals.

3.8.2.B In conventional indexes arranged by subject headings

Such printed indexes include al-Fihrist (الفهرست) and other similar manual indexes which cover the contents of Arabic specialized and academic journals. Titles of articles appear in every single record included in such indexes. Chapter two of this thesis provides detailed information on such indexes and about the bibliographical descriptions included in their records.

3.8.2.C General and specialized bibliographies

The depth of indexing in such Arabic services varies widely. The majority of these publications are usually not up-to-date and their subject headings are broad rather than specific. Because the number of articles under such subject heading is considerable, dependence has to be on titles of articles in deciding whether or not to go and find the full articles. Examples of such bibliographies are: Bibliography of the Arabic Unity (الوحة العربية) which is published by the Centre for Arab Unity Studies in Beirut, and Bibliography of the Islamic Maghrebin (الغرب الإسلامي) which is published by King ‘Abd al-‘Aziz Foundation for Islamic and Humanities Studies in Casablanca. A sample of the latter is shown in
3.8.2.D Titles indexes

These are services in which the first word of the title of article is likely to be the filing word, and thus the most important element and the only access point. Individual printed indexes for Arabic periodicals often have two sections: the first is usually called subject index, but in fact it is a straightforward title index, and the second is the author index. A sample of such an index taken from Message of the Library Cumulative Index for the period between 1965 - 1985 is shown in figure 3.20.

3.8.2.E In machine-readable databases

Mechanized information retrieval systems which allow searching in Arabic script were not made available in the Arab World until recently. It was only in 1982 that a breakthrough was done when ASMO 449 (Arab Organization for Standardization and Metrology) became the standard coding system for the Arabic language. Consequently the business of Arabizing information retrieval systems has started and searching by keywords in Arabic titles and other fields has become available. Ashoor (1989a) reports three Arabized automated retrieval systems that are extensively used in libraries in the Arab World. These are: DOBIS/LIBIS, MINISIS, and STAIRS. Also section 7.1 in this thesis quotes several studies which report Arabic script compatible retrieval systems used for books, journal articles as well as other Arabic printed materials.
علوم التربية

157
أبوحجيب، موان سليم
悩み أصل الفقه الإسلامي في الفكر النسائي، موان سليم أبوحجب، [شنم محمد سعيد رضى] - [دم] : الناد。
المجلة العامة 1987 - 528 ص: نبات مه. بالألون، 24 م.
تينبيرجا، ص. 90-524

158
المجلة العربية، المجلة العربية، ديبورات المعرفة/ التأثير الوسيط/ الأسس الإسلامية، من حضور، выполнено، في الدين.
أبوحجيب، موان:

159
المجلة العربية، المجلة العربية، ديبورات المعرفة/ التأثير الوسيط/ الأسس الإسلامية، من حضور، выполнено، في الدين.
أبوحجيب، موان:

160
المجلة العربية، المجلة العربية، ديبورات المعرفة/ التأثير الوسيط/ الأسس الإسلامية، من حضور، выполнено، في الدين.
أبوحجيب، موان:

161
المجلة العربية، المجلة العربية، ديبورات المعرفة/ التأثير الوسيط/ الأسس الإسلامية، من حضور، выполнено، في الدين.
أبوحجيب، موان:

Figure 3.19 : sample of the Bibliography of the Islamic Maghrebin
كشاف العناوين

الرقم
03
72
868
105
433
361
187
186
170
339
9
9
142
555
91
91
439
443
370
85
86
367
332
366

اتحاد جمعيات المكتبات العامي.
الاتحاد الدولي للنونيق: دوره ونشاطاته.
أثر استخدام التقنيات السمعية المصرية.
أثر دائرة المعارف البريطاني في الثقافة العالمية.
الاحترام في الدراسات الإسلامية.
احصاءات عن المكتبات المتخصصة في الأردن لعام 1997.
اختيار الكتب في المكتبات العامة.
اختيار الكتب في مكتبة الأطفال.
اختيار مداخل المؤلفين العرب وقواعد الفهرسة الوصفية للمحتوى.
إدارة النونيق والمعلومات في المنطقة العربية للعلوم الادارية.
ذب الأطفال في الأردن.
ذب الأطفال. ووضعوا.
ذب الأطفال ومكناتهم.
إزالة النشر في الوطن العربي.
السيلة المرجعية وأنواعها في المكتبة العامة.
الاستجابات لوسائل الاتصالات.
استخدام الحاسبات الآلية في أعمال المكتبات والتنويم.
استخدام الكمبيوتر في مكتبة وزارة التخطيط.
الاستقرار الوظيفي عند المكتبيين.
استطاعة حول خدمة التصوير في مكتبة الجامعة الأردنية.
استطاعة عن بعض مراكز المعايير في عمل.
استطاعة عن المكتبة الإسلامية في الزرقاء.
استطاعة عن مكتبة بلدية طلمس العامة.
استطاعة عن مكتبة البنك المركزي الأردني.

رسالة للمكتبة.

Figure 3.20 : sample of the Message of the Library’s title index
3.9 Advantages and drawbacks of title keyword indexing

As with any other indexing technique, the title keyword indexing of articles has its advantages and drawbacks. The English writings in the field of indexing is saturated with studies which have approached these issues ever since its introduction to the contributors of the International Conference of Scientific Information held in Washington in 1958. The following two sections represent a condensed review of what has been said about the title keyword indexing of articles in English. Whether the same features apply to titles of Arabic articles is for the following chapters to reveal.

3.9.1 Advantages

a - A main advantage which has been claimed for printed indexes using title keywords of periodical articles is the promptness with which they can be produced. Actually, as a current awareness indexing tool, they provide a quick means to make their users aware of all new articles which have been controlled.

b - Because of the mechanical method of preparation of title keyword printed indexes, a larger quantity of information can be processed within a given time, and more access points are provided than would have been practical by conventional means. The number of access points depends on two elements:

i - the length of the article title and the allotted space provided by the computer.

ii - the length of the keywords in the title itself.

The provision of multiple access points improves the probability that a particular article will be retrieved.
c - A title keyword index costs much less to generate than any conventional subject index.

d - The use of a computer to generate title keyword printed indexes eliminates indexing delays.

e - Title keyword indexing processed by means of computer is unlike conventional indexing methods: it can be conducted with no or little human intervention.

f - The production of a printed title keyword index is indeed simple and very straightforward. A considerable number of title keyword indexing techniques have been written to be used with a variety of computer configurations, which offer a wide range of input and output options.

g - Title keyword index entries which are in machine readable format can be used for other purposes such as the preparation of special bibliographies and accession lists and SDI (Selective Dissemination of Information) services.

h - In title keyword indexing, the indexing terms are always current and are the actual words of authors.

To these features, many can be added which relate to the type of title index produced and to the type of computer hardware and software used in the generation process.

3.9.2 **Drawbacks**

a - The most common complaint against the title keyword indexing techniques is the lack of terminology control which brings about many serious problems of synonyms, near-synonyms, homographs and the inclusion of many irrelevant entry points. However, these problems have been alleviated with the use of a thesaurus in machine-readable form and the
use of cross-references as well.

b - Since the article title is the only list from which the keywords will be extracted, the need for informative and representative titles imposes itself.

c - The information content of an article is not always entirely discernable from the title.

d - Objections have been claimed with respect to the difficulties encountered in reading from a title keyword index because of the size of the type face. These objections have now been alleviated and title keyword indexes can be prepared with considerably improved readability, as well as format and type face.

e - The existence of irrelevant entry points in the indexing window places a burden on the users of title keyword indexes, who frequently need to search through a number of irrelevant entry points before catching pertinent articles.

f - The limited space allotted to the index line entry, especially with the KWIC index types, causes mutilation to all article titles which exceed that space. Thus, the meaning of truncated titles may not be clear, or may be misleading.

3.10 Conclusion

In this chapter almost all types of title keyword indexing techniques which cover the contents of English periodical articles were discussed, as well as the functions of Arabic article titles in the current indexing tools.

As a matter of fact, the absence of title keyword indexing techniques from the Arab world does not necessarily mean that titles of Arabic periodical articles do not have other functions in the current Arabic indexing tools. It has been shown in the course of this chapter that titles have other functions to
play alongside the bibliographical information of Arabic periodical articles. This absence does not also necessarily mean that titles of Arabic periodical articles are inadequate to be used for such indexing techniques.

Having seen that title keyword indexing is largely used in the West for both scientific and non-scientific periodical articles whether in printed or machine readable form, there is no obstacle to the examination of Arabic article titles to be used for producing printed indexing by means of computers. Especially, as shown in chapter two, since the current Arabic indexing tools appear to be insufficient and always in arrears.

To investigate whether this type of indexing is applicable for Arabic periodical articles is the function of the following chapters through the examination of all other related issues. The following chapters of part two deal with issues of Arabic titles' informativeness and representation of their accompanying periodical articles.
chapter four

The information content of titles of Arabic articles and comparison with English titles

This chapter investigates the informativeness of titles of Arabic periodical articles through the examination of their information content. Since titles of English periodicals have been reported in the literature as informative, the results obtained from examining Arabic titles are compared to those of a new study for the content of English titles. This chapter represents a partial answer towards the possibility of using keywords of Arabic article titles to produce printed indexes by means of computer.
4.1 Literature survey

This section reviews English and Arabic studies which examined the information content of titles. The concept of information content as used in this examination, is the counting of words, numbers and all types of written characters which appear in the selected titles. However, specialists distinguished between these entities when they conducted their studies. This distinction is clarified during the course of this chapter.

4.1.1 English titles

The prior studies which investigated the informational value of titles of English periodical articles have focused upon the variation of the titles content with time and disciplines in scientific literature. Some of these also investigated titles of articles from the social sciences and humanities.

The first author to measure the information content of titles by counting their substantive words and suppressing the nonsubstantive ones was Jaques J. Tocatlian. In his research, Tocatlian (1970) did not cite any prior studies, but was cited by several studies of the subsequent works. None of these referred to research prior to that of Tocatlian. Table 4.1 shows the works cited by the subsequent papers. No other common references were found among these studies.
<table>
<thead>
<tr>
<th>Y.P.*</th>
<th>Author</th>
<th>N.R.**</th>
<th>Cited authors</th>
<th>Cited</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970</td>
<td>Tocatlian</td>
<td>--</td>
<td>--</td>
<td>6</td>
</tr>
<tr>
<td>1975</td>
<td>Bird &amp; Knight</td>
<td>4</td>
<td>T</td>
<td>5</td>
</tr>
<tr>
<td>1977</td>
<td>Buxton &amp; Meado.</td>
<td>5</td>
<td>T, BK</td>
<td>5</td>
</tr>
<tr>
<td>1981</td>
<td>Balog</td>
<td>6</td>
<td>T, BK, BM</td>
<td>1</td>
</tr>
<tr>
<td>1982</td>
<td>Diodato</td>
<td>7</td>
<td>BM</td>
<td>1</td>
</tr>
<tr>
<td>1984</td>
<td>Diener</td>
<td>4</td>
<td>T, BK, BM, BA</td>
<td>1</td>
</tr>
<tr>
<td>1987</td>
<td>Buxton</td>
<td>5</td>
<td>BM, D, T, BK</td>
<td>1</td>
</tr>
<tr>
<td>1991</td>
<td>White &amp; Hernan.</td>
<td>11</td>
<td>T, BK, BM, DI, B</td>
<td>--</td>
</tr>
</tbody>
</table>

Table 4.1: Composite of the prior research authors. *= year of publication; **= number of references; T: Tocatlian, BK: Bird and Knight, BM: Buxton and Meadows, BA: Balog, D: Diodato, DI: Diener, B: Buxton.
Despite the differences in the size of samples, and in the number of subjects or sources, the authors of the above mentioned studies had in common one objective. This was to find out whether titles of research papers are becoming more informative since the application of title keyword indexing techniques and the proliferation of machine-readable databases which can be searched by title words.

Tocatlian (1970), after investigating samples of chemical literature from the years 1948, 1958, and 1968 concluded that "the statistical significance of the difference we measured is such that we can say with confidence: yes, titles of chemical papers are becoming more informative".

In order to satisfy his first hypothesis that "titles of chemical articles are becoming more informative since the advent of KWIC indexes", he suggested a set of five measurements which can be used for the study of the content of titles. One relevant measurement for this chapter is that which investigates the informativeness of titles by counting their substantive words and rejecting the others which "convey little or no information by themselves, such as articles, prepositions, conjunctions, pronouns, and auxiliary verbs".

From a sample of 300 article titles taken from 10 chemical journals gathered during the years 1948, 1958, and 1968 on a basis of 10 titles for each of the three years for each of the 10 journals (at 100 titles per year), he calculated the mean numbers of substantive words per title by years as shown in table 4.2. Then, in order to assess the growth in the information content of titles of chemical articles, he calculated the differences between the yearly means and got the results shown in table 4.3.
<table>
<thead>
<tr>
<th>Year</th>
<th>Substantive words per title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1948</td>
<td>5.57</td>
</tr>
<tr>
<td>1958</td>
<td>5.46</td>
</tr>
<tr>
<td>1968</td>
<td>6.77</td>
</tr>
</tbody>
</table>

Table 4.2: the number of substantive words found in titles of chemical articles during the selected years - Tocatlian (1970).

<table>
<thead>
<tr>
<th>Year</th>
<th>Differences in means</th>
</tr>
</thead>
<tbody>
<tr>
<td>1948 - 1958</td>
<td>-0.11</td>
</tr>
<tr>
<td>1958 - 1968</td>
<td>1.31</td>
</tr>
<tr>
<td>1948 - 1968</td>
<td>1.20</td>
</tr>
</tbody>
</table>

Table 4.3: the difference in means found between the selected years - Tocatlian (1970).
Tocatlian (1970) observed that a statistically significant difference was found between the years 1958 and 1968, whereas no such difference was detected between 1948 and 1958.

The first work which confirmed Tocatlian's findings came after five years, when Bird and Knight (1975) investigated the change in the information content of titles with time in four scientific journals: *Nature*; *Journal of Clinical Endocrinology and Metabolism*; *Analytical Chemistry*; and *Chemical Abstracts* (analytical section only) during the years 1954, 1964, and 1974 by counting the total number of words per title and for *Nature* only by counting the number of substantive words as well (as determined by the 650 common words on the Permuterm Index stoplist). They concluded that "titles in scientific journals are becoming longer and hence more informative", and that the average of the increase in title length for the four sources over the 20 years is 2.4 words (a difference which is significant at the 5 percent level by T test).

To measure the change over time Bird and Knight (1975) followed the technique adopted by Tocatlian (1970) and found a rapid increase during the decade between 1954 and 1964, but only a slight increase during the decade between 1964 and 1974. Table 4.4 shows the average title length in each of the three years for the four selected journals. However, these results may be put in two tables 4.5 and 4.6, since the investigated journals were from two disciplines: from science in general represented by *Nature* in table 4.5, and from the chemical literature represented by the other three journals in table 4.6.
<table>
<thead>
<tr>
<th>Year</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1954</td>
<td>8.70</td>
<td>3.93</td>
</tr>
<tr>
<td>1964</td>
<td>10.0</td>
<td>4.21</td>
</tr>
<tr>
<td>1974</td>
<td>11.0</td>
<td>4.27</td>
</tr>
</tbody>
</table>

Table 4.4: average of titles length for the four selected journals - Bird and Knight (1975).

<table>
<thead>
<tr>
<th>Year</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1954</td>
<td>7.70</td>
<td>3.44</td>
</tr>
<tr>
<td>1964</td>
<td>8.80</td>
<td>4.01</td>
</tr>
<tr>
<td>1974</td>
<td>9.00</td>
<td>3.43</td>
</tr>
</tbody>
</table>

Table 4.5: average of titles length in Nature - Bird and Knight (1975).

<table>
<thead>
<tr>
<th>Year</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1954</td>
<td>9.00</td>
<td>4.08</td>
</tr>
<tr>
<td>1964</td>
<td>10.4</td>
<td>4.28</td>
</tr>
<tr>
<td>1974</td>
<td>11.8</td>
<td>4.51</td>
</tr>
</tbody>
</table>

Table 4.6: average of titles length in chemical literature - Bird and Knight (1975).
Bird and Knight (1975) believed that there were two reasons for the trend towards more informative titles:

a - An increased awareness by authors of the importance of longer and more informative titles as carriers of information, especially in disciplines where research is more intensive.

b - The effect of increasing use of title-based indexes and title-word searchable databases, and of lists of CC (Current Contents).

One difference between this study and that of Tocatlian (1970) is that the latter counted the substantive words to conclude that the titles were becoming more informative, whereas Bird and Knight (1975) relied on the increasing number of all words per title to reach their conclusion. This result needed more confirmation because they did not attempt to discover the kinds of words which were represented by the increase in the number of all words per title over the selected periods.

Buxton and Meadows (1977) examined the content of article titles from a variety of subjects and found a general trend towards increasing numbers of substantive words per title. They concluded, for the first time in a paper which handles scientific and non-scientific disciplines, that "on the basis of the number of information words they contain, the titles of research papers in physics, history, psychology and to a somewhat lesser extent other social sciences do not seem to fall far short of chemistry and the life sciences in their suitability for retrieval".

Their methodology involved investigating 100 article titles each taken from eleven English language journals during the time periods of 1947, 1962, and 1973, with some journals being sampled over greater than a one-year period to get a sufficient sample size. They measured the information content of titles of research papers by counting the number of words per title; the
number of substantive words per title, as determined by an unpublished stoplist of 230 words; and then by getting the proportion of substantive words per title. The journals covered were in the fields of chemistry, economics, history, life science, philosophy, physics, psychology, and sociology. Each field was sampled by one journal except chemistry and life science which were sampled respectively by three and two.

Apart from the general trend towards more informative titles which has occurred over a wide range of subjects (philosophy being the only exception), the results of this study indicated a significant increase in the number of substantive words in the fields of chemistry, life science, and in economics during the period 1962 - 1973. They also showed a significant increase in history throughout 1947 - 1962. For a detailed table of the results for each journal see the original paper.

Buxton and Meadows (1977) questioned the effect of the introduction of the B.A.S.I.C. index in 1959 and Chemical Titles in 1960 on the increase in information content of chemical and biological titles. They said that this introduction cannot provide an explanation for the trends towards more informative titles which they found in the social science and arts subjects, and on chemical titles which showed significant increase in the period 1947 - 1962. Because "in the social science and arts subjects new techniques and new aspects of study are comparatively much rarer, so that their influence on the information content of titles is much less". Their findings are reorganized in table 4.7 for their study on scientific journals, and in table 4.8 for the non-scientific journals.
Table 4.7: Buxton and Meadows (1977) results for titles in scientific journals. # = all words per title; != standard deviation; * = substantive words per title.

<table>
<thead>
<tr>
<th>Year</th>
<th>Wd.Tit.#</th>
<th>S.D.!</th>
<th>Sb.Wd.Tit*</th>
<th>S.D.!</th>
</tr>
</thead>
<tbody>
<tr>
<td>1947</td>
<td>9.72</td>
<td>4.22</td>
<td>5.71</td>
<td>1.29</td>
</tr>
<tr>
<td>1962</td>
<td>9.22</td>
<td>2.01</td>
<td>5.95</td>
<td>0.55</td>
</tr>
<tr>
<td>1973</td>
<td>11.21</td>
<td>3.26</td>
<td>7.37</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Table 4.8: Buxton and Meadows (1977) results for titles in non-scientific journals. # = all words per title; != standard deviation; * = substantive words per title.

<table>
<thead>
<tr>
<th>Year</th>
<th>Wd.Tit.#</th>
<th>S.D.!</th>
<th>Sb.Wd.Tit*</th>
<th>S.D.!</th>
</tr>
</thead>
<tbody>
<tr>
<td>1947</td>
<td>7.19</td>
<td>2.27</td>
<td>4.24</td>
<td>0.77</td>
</tr>
<tr>
<td>1962</td>
<td>7.65</td>
<td>2.46</td>
<td>4.66</td>
<td>1.04</td>
</tr>
<tr>
<td>1973</td>
<td>8.17</td>
<td>3.64</td>
<td>4.95</td>
<td>1.28</td>
</tr>
</tbody>
</table>
Finally, Buxton and Meadows (1977) stated that as far as retrieval is concerned "for a particular title, the number of substantive words is not simply related to its value in retrieval either as regards recall or precision. However, it seems reasonable to suppose that in general a longer title will contain more words related to the subject matter of the paper, and will be of more use as a basis for retrieval ".

Balog (1981), after investigating the content of article titles in the field of agriculture as represented in the *New Zealand Journal of Agricultural Research* concluded that "titles of papers in this agricultural journal are becoming longer but they are also becoming more informative as judged by the number of substantive words per title ". He followed the lead of Buxton and Meadows (1977) by counting the number of all words per title; the substantive words per title; and then by getting the proportion of substantive words per title. However, he had added new aspects to his study when he classified the examined papers into singular author vs. multiple author, and series vs. non-series articles.

Balog's population was the papers published in *New Zealand Journal of Agricultural Research* from 1958 (volume 1) to 1978 (volume 21). However, data were reported for only six volumes as shown in his table of results (see table 4.9).

Balog's results (1981) are in line with those of the previous works. However, two more main observations emerged from his study; the first is that groups of authors are better at producing informative titles than are single authors, and the second is that there must be a limit to the number of words in titles of scientific papers and some of these have already reached the 12 words title recommended by O'Connor and Woodford (1977, p.47). Balog suggested that single authors seek the opinions of their colleagues in the
composition of the titles of their papers.

<table>
<thead>
<tr>
<th>Year</th>
<th>V.</th>
<th>Wd.Tit.*</th>
<th>Sb.Wd.Tit.**</th>
<th>Pr.Sb.Wd.Tit.#</th>
</tr>
</thead>
<tbody>
<tr>
<td>1958</td>
<td>1</td>
<td>12.70</td>
<td>7.07</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>13.86</td>
<td>8.46</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>12.70</td>
<td>7.66</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>13.68</td>
<td>8.47</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>12.96</td>
<td>8.37</td>
<td>0.64</td>
</tr>
<tr>
<td>1978</td>
<td>21</td>
<td>14.84</td>
<td>9.78</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Table 4.9: Balog’s (1981) results taken from his study. V. = volume number; * = average of total words per title; ** = average of substantive words per title; # = proportion of substantive words per title.

Diodato (1982) investigated article titles from five journals, four of them were previously studied by Buxton and Meadows (1977). His population was a sample of 50 titles each taken from the five selected journals which cover subjects in chemistry, economics, history, philosophy and mathematics. To provide some consistency between the method he used and the method used by Buxton and Meadows, Diodato examined for a "given journal in a given decade the first year examined by the earlier workers" and assigned the two dates 1962 and 1973 as investigated years for his added fifth journal.

This study supports the basic findings of Buxton and Meadows (1977) that there was an increase in the number of keywords per title in chemistry and economics and a decrease in philosophy over the period 1962 - 1973. However, if Diodato’s results are considered solely, the study shows a
significant change (an increase) in the mean number of title keywords over time in chemistry, economics and mathematics. There is no significant change in the humanities fields, history and philosophy (see table 4.10).

<table>
<thead>
<tr>
<th>Subject</th>
<th>Mean and (S.D.*) of substantive words per title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemistry (1st decade)</td>
<td>6.84 (2.41)</td>
</tr>
<tr>
<td>Chemistry (2nd decade)</td>
<td>7.88 (2.51)</td>
</tr>
<tr>
<td>Economics (1st decade)</td>
<td>4.38 (2.00)</td>
</tr>
<tr>
<td>Economics (2nd decade)</td>
<td>5.60 (1.71)</td>
</tr>
<tr>
<td>History (1st decade)</td>
<td>5.44 (1.95)</td>
</tr>
<tr>
<td>History (2nd decade)</td>
<td>5.12 (1.98)</td>
</tr>
<tr>
<td>Mathematics (1st decade)</td>
<td>3.90 (1.37)</td>
</tr>
<tr>
<td>Mathematics (2nd decade)</td>
<td>4.62 (1.34)</td>
</tr>
<tr>
<td>Philosophy (1st decade)</td>
<td>3.28 (1.64)</td>
</tr>
<tr>
<td>Philosophy (2nd decade)</td>
<td>3.06 (1.44)</td>
</tr>
</tbody>
</table>

Table 4.10: Diodato's (1982) results taken from his study. *= standard deviation.

Diener (1984a, p.48) investigated the informational dynamics of journal article titles by analysing the change of the informational value of titles over time in four selected English language journals. He claimed that in his research the prior studies of information content variation of journal article titles over time were extended in two ways:

a - An evenly-distributed, longitudinal sample of journal article title is selected covering the thirty years from 1951 through 1980 and to assess the
pattern of change over time a linear regression analysis technique was used.

b - The information content measures followed in the previous studies were augmented by a measure of what he called "the relational value of titles". To measure the relational value, each title is transformed from its linguistic form to a structural form through relational analysis. Then "the number of relations in the structural image is counted". According to Diener (1984a, p.22), "this technique has evolved out of the work by Jason Farradane on relational indexing for document retrieval".

As for the journals selected for sampling, they were chosen using four criteria.

a - The journal had to have been in continuous publication throughout the chosen period.

b - The journal had to be either a physical science journal or a social science journal.

c - The journal had to be either a primary or a secondary (review) journal.

d - The journal had to have a relatively high rank by impact factor reported by ISI (the Institute for Scientific Information). "impact factor is a measure of the relative citedness of a journal. It is computed by dividing the number of times a journal is cited in the literature over a prior two-year period by the number of articles published in the current year's volume".

In this research, four titles per year were randomly sampled, so that each journal was represented by 120 titles and from each of these titles the informational value was measured as:

a - The total number of words per title.

b - The number of substantive words per title.

c - The proportion of substantive words per title.
The relational value, or the number of relations per title.

After analysing the titles Diener (1984a, p.73) concluded that the informational dynamics of journal article titles, assessed as a macro-phenomenon has remained relatively stable and consistent throughout the literature over the years from 1951 to 1980 and "those who rely upon titles, such as information seekers... should not expect any significant improvement in the usefulness of titles to inform about the contents of the document". However, the informational dynamics of journal article titles assessed as a micro-phenomenon "provides a means of investigating the similarities among, and differences between, various subsets of the literature" such as the study of physical science and the social science literatures as well as the primary and review literatures. Diener (1984a, p.74) reported that "no significant differences in the number of words, keywords, or rates of change existed between either pair".

Later on in the same year, Diener (1984b) carried out another research on the same topic for the same period. He added to his previous sample size, the sum of 1638 article titles taken from six various journals specializing in library science and another 1848 titles from miscellaneous journals. These titles were also analysed by measuring the total number of words per title, the number of substantive words per title, and the proportion of substantive words per title.

In this research, Diener (1984b) confirmed his previous results and stated that "the titles of journal articles and particularly the substantive words in titles tend to behave in regular, linear ways". He also suggested that titles in physical science tend to contain 10% more keywords than social science titles. This appears to confirm Buxton and Meadow's conclusion which was suggested in their paper in 1977, but the low values suggest not much real
Buxton (1987) examined the content of titles taken from the same eleven English language journals used in his earlier work (Buxton and Meadows, 1977). The samples of 100 titles each were taken from these eleven journals for the year 1984. Following the lead of his earlier work, Buxton (1987) concluded that "the length of titles in the natural science journals sampled seems to have stopped increasing since 1973, and the average is now ten to thirteen words ". He believed that the increase in the length of titles of scientific papers has levelled off, and that a "saturation point " has been reached. Moreover, he suggested that any "further increase in length would make titles unwieldy ".

Based on the fact that "every field has its special language, and if the field grows in complexity, the vocabulary of its special language enlarges ", White and Hernandez (1991) carried out an inter-disciplinary study to investigate whether the new field of counselling has developed like other older fields as mirrored in article titles of its principal periodicals. For this investigation, two principal sets and one comparison set of article titles were selected. The first set comprised nine journals in continuous existence since 1923 chosen from the fields of education, psychology, social work, sociology, and economics. The second set consisted of five periodicals representing the field of counselling and one comparator journal, the Journal of Social Psychology. Finally, to provide comparison data from periodicals with a relatively short life span, three recently published journals constituted the third set.

In order to test the expectation that the fields represented by the examined article titles were becoming more complex over time, White and Hernandez (1991) used the multiple regression and analysis of variance
techniques. Their findings were that the examined "fields generally have been becoming more complex, including the development of more complex research procedures, and greater specificity and elaboration of field-associated theories".

Having seen the results reported by these studies, one can say that there is a thread of consistency in the way of counting the words in English titles as well as in showing that titles are becoming more informative, although a disagreement on reasons behind this trend was observed. Scholars who related the increase in the number of keywords in English titles to the introduction of KWIC indexing techniques were right in their suggestion, and scholars who questioned it were right as well. But to know which of these groups was more right, studies of English titles for the period prior to 1958 are needed.

In fact, the disagreement on the possible causes for the trend towards longer and more informative titles which resulted from the above studies, might be cleared if the contents of article titles in a language other than English are investigated. To achieve a reasonable answer, the condition of unfamiliarity of authors and users of this language with title keyword indexing techniques has to be fulfilled.

One language which meets these conditions is the Arabic language. But before getting into this work a word is due here, that is the study of informativeness of Arabic article titles which forms the core of this chapter was not chosen for investigation in order to clear the reported disagreement on English titles. It was suggested that the results which will be obtained after carrying out this new study might help in clearing it up. Of course, the characteristics of each of these two languages and their effects on the construction of article titles need to be discussed.
4.1.2 Arabic titles

The issue of using title word indexing techniques as a service for indexing the content of Arabic periodical articles does not seem to have been thoroughly and properly investigated, nor has there been a study of their representation of the contents of the corresponding articles. All computer searches, which were conducted, and consultations of the available indexing and abstracting services, have produced nothing relevant to this issue.

The only relevant work which approached the information content of Arabic article titles was carried out by H.M. Kasem in 1978 at University College London. Kasem studied the Arabic language in specialist information systems. In a small part of a chapter of a Ph. D. thesis, Kasem (1978, p.197) studied a sample of 434 titles drawn randomly from Arabic journal literature in linguistics, and found that the average number of main entries per title was 3.5; then he translated these titles into English in order to compare the efficiency of computer retrieval between Arabic and English titles in one specialist information system. Later, Kasem (1985) made a suggestion for Arab researchers to investigate the possibility of using title word indexing techniques in covering the content of Arabic periodicals.

Two years later, al-'Atram (1987) published an Arabic article on the use of natural language in information retrieval and its application to the Arabic language. At the end of this article, al-'Atram mentioned that he was going to undertake research funded by the King 'Abd al-'Aziz City for Science and Technology to examine the characteristics of titles of Arabic articles in some Arabic journals. However, al-'Atram never published his study; no part of his work has been reported, cited, or published in the current journals. Afterwards, al-Suwaydan (1993) approached the issue of subject retrieval
using Arabic title words and employed 187 titles taken from 18 Arabic periodicals in order to investigate the usefulness of Arabic titles in this domain. Al-Suwaydan did not examine the informativeness of Arabic titles and did not investigate their representation. He only mentioned some of the problems that are encountered in the process of retrieval by depending on title keywords. It is relevant to mention here that al-Suwaydan did not cite any previous related works and missed out the results of this chapter which were published earlier to his work in 1991 as mentioned in the acknowledgements of this thesis. At the end of his work, al-Suwaydan emphasised the importance of this issue and mentioned that he is willing to expand the aspect of his study in the future.

4.2 A study of the information content of Arabic titles

4.2.1 Objectives of the study

The aim of this study is to provide an answer to the question of whether Arabic titles are informative. This answer alone is not satisfactory to tell whether Arabic titles can be relied on as a basis for title word indexing techniques. Another study has to be undertaken; that is the investigation of the representativeness of these titles, since an informative title is not necessarily representative of its article. This difference was not given any attention in some of the studies of English titles which were mentioned earlier in this chapter. Authors counted the keywords in titles to find out whether titles are becoming more informative but did not determine if they were more representative.

4.2.2 Method and samples

In this new study of Arabic titles, the method used by Buxton and
Meadows (1977) was followed with one notable difference, that is only one
date period was taken. As mentioned earlier, no thorough work has been done
on Arabic titles before, and indexing by each keyword in titles has not been
yet introduced in the Arab countries. So it is not possible to study its effect
on title content changes by authors between two dates.

This study investigates titles in at least two periodicals in each scientific
subject and more than three in each non-scientific subject. This is because one
periodical does not necessarily represent the titles in one subject. Also,
sticking to the titles of one journal will restrict the study to one journal’s
article titles in which case, the titles may be affected by the comments of
referees and editors of that journal, and make it untypical of titles in that
field.

In order to compare across the various selected subjects and to examine
the informativeness of titles it was necessary to adopt some conventions about
what was counted as a keyword and what was defined as a stopword or
non-keyword. In this study, any single or set of characters which occurred
between two spaces was considered to be a word. Stopwords were defined as
those which carry no useful information by themselves, such as prepositions,
articles, conjunctions, pronouns, introductory prefixes (inseparable
prepositions), as well as single letters. Multiword names are counted
separately; chemical formulae are counted as single words; numbers e.g.
1989, with following symbols e.g. 1300 AH (Islamic Calendar) or 1200 AD
are counted as single words as well; hyphenated words are also counted as
single; dates and abbreviations are counted as single words. A hyphen which
separates two dates is suppressed and each date is counted as one word.
Subtitles are included; English or Latin words in the Arabic title are counted.

To carry out this study, samples of 50 titles each were taken from 16
different subjects in both scientific fields (covering the years 1980-1988) and non-scientific (covering the years 1980 - 1985). Titles of non-scientific subjects were collected from the Arabic indexing service al-Fihrist (الفهرست), but before considering the dependence on titles from this tool, two steps were taken to examine the accuracy of titles cited.

a - A sample of 50 titles from different subjects was tested to ensure that titles in al-Fihrist bibliographic records are printed as they are in their original corresponding articles.

b - In verbal communications, a senior indexer in al-Fihrist has confirmed their policy of copying the titles of indexed articles.

Having been collected, titles were then exposed to a reliability test which showed agreement between two readers on the classification of titles into subject fields.

As far as scientific titles are concerned they were taken from Arabic periodicals which publish articles in Arabic as well as in English, with the provision of Arabic titles and abstracts when the articles are in English. Only articles by Arab authors were considered. Scientific journals published completely in Arabic are few, and inaccessible because of the absence of an index to cover their contents. This is why the date period for scientific samples had to be expanded, and titles had to be taken from the periodicals themselves. The selected scientific and non-scientific Arabic periodicals that were used for sampling in this study are shown in appendix 4.1.

It is worth mentioning here that during the counting of keywords in both scientific and non-scientific titles, the process came across many substantive and stopwords with introductory prepositions. In such cases and because this chapter is examining the information content of Arabic article titles as a part of a study which investigates the viability of these titles for the production of
keyword printed indexes, introductory prefixes such as inseparable prepositions were separated from the keywords. For instance, with relevance to the definition of a word, the following title:

التطور التاريخي للمكتبات في ليبيا

( the historical development of libraries in Libya) consists of five words, but if the inseparable preposition ل "L", which matches the word " of " in English, is separated from the word المكتبات " (of libraries) the number will increase to six and the title will be written as the following during the counting process

التطور التاريخي للمكتبات في ليبيا

Inseparable prepositions had to be separated, because otherwise the alphabetical order of a single Arabic keyword with different introductory prepositions would be scattered throughout a keyword index. Whereas the skipping of suffixed letters will not disrupt the alphabetical sequence; e.g. see table 4.12 which takes the keyword فهرسة (fahrasah, cataloguing) as an example. The first letter in this keyword ف (F) is sequenced 27th in the Arabic alphabet. But as we notice in this table, the added prepositions, if left, will scatter the same keyword in different places throughout the keyword index.
Table (4.11): Sample of an Arabic keyword with three various inseparable prepositions. *: Sequence number in the Arabic alphabet when it is sorted according to ASMO 449 (Arab Organization for Standardization and Metrology).

4.2.3 Arabic stopwords list

The dearth of published Arabic stopwords lists has compelled this study to rely on an existing English one, which can do the job with the existence of a good translation. The Arabic specialized stopword list, used by Kasem (1978) when he studied Arabic titles in linguistics, is not included in his thesis. Otherwise, the two published general Arabic stopword lists which could be located did not include all sorts of stopwords. The first, reported by Kaddim et al. (1988) in their paper on computerized text analysis based on transition phenomena, comprises 17 trivial words and four special symbols. The second stopword list consists of about 130 stopwords and appeared in a book in Arabic by al-Suwayni' (1988, p.50) about permuted indexes and information retrieval in Arabic. Therefore, to accomplish this study, an existing English stopword list of about 230 words (including inflexional variants) compiled by Buxton and Meadows (1977) was adopted, and enhanced with other uncommon stopwords found in the two mentioned Arabic lists or missed out.
by all of them.

Besides the inadequacy of the reported Arabic stopword lists, there were some other considerable reasons behind the adoption of Buxton and Meadows' s stopword list as a starting point. These reasons were the following:

a - Establishing a stopword list is not a straightforward process, because words deemed non-significant in one subject might be useful in another; e.g. "road" should probably be suppressed in library science, but not in civil engineering; "assets" will be useful in accounting but not in sociology; and "geography" should be stopped in a geography index but not in agriculture.

b - This stopword list is confined to prepositions, pronouns and auxiliary verbs, together with general terms such as "aspects" and "application" which seemed safer in this study which is dealing with multidisciplinary topics.

c - Among the results of the contents of English article titles examination, shown in section 4.3.3, four subjects were taken from Buxton and Meadows's (1977) study in which they depended on this stopword list. The examination of English titles was carried out for later comparison with that of Arabic titles. The four subjects were those of chemistry, physics, medicine and engineering.

The full enhanced stopword list which was used in this study is shown in appendix 4.2. It consists of approximately 300 Arabic stopwords. As for the possibility of extending this stopword list to cover a probable shortage, Kasem (1978, p.197) revealed, after he extended the stopword list which he used in his study of Arabic titles in linguistics from 61 to 129 stopwords, the decrease in the average number of entries from titles is 8%. This is a ratio too
small to affect the conclusions.

4.2.4 Results

The results of the study of the information content of titles of Arabic periodical articles are given in table 4.12 for the scientific titles, and in table 4.13 for the non-scientific titles. These tables provide the mean and standard deviation numbers for titles of the examined subjects.

<table>
<thead>
<tr>
<th>Subject</th>
<th>All words</th>
<th>Substantive words</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.*</td>
</tr>
<tr>
<td>Agriculture</td>
<td>13.64</td>
<td>4.32</td>
</tr>
<tr>
<td>Chemistry</td>
<td>12.62</td>
<td>4.95</td>
</tr>
<tr>
<td>Engineering</td>
<td>11.00</td>
<td>3.89</td>
</tr>
<tr>
<td>Medicine</td>
<td>10.84</td>
<td>5.21</td>
</tr>
<tr>
<td>Physics</td>
<td>10.44</td>
<td>4.47</td>
</tr>
</tbody>
</table>

Table 4.12: the mean and standard deviation (S.D.*) numbers of contents of Arabic scientific article titles in five disciplines.
<table>
<thead>
<tr>
<th>Subject</th>
<th>All words</th>
<th>Substantive words</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.*</td>
</tr>
<tr>
<td>Economics</td>
<td>9.40</td>
<td>3.60</td>
</tr>
<tr>
<td>Education</td>
<td>9.88</td>
<td>4.10</td>
</tr>
<tr>
<td>Geography</td>
<td>8.00</td>
<td>3.70</td>
</tr>
<tr>
<td>History</td>
<td>9.72</td>
<td>3.95</td>
</tr>
<tr>
<td>Law</td>
<td>9.22</td>
<td>4.50</td>
</tr>
<tr>
<td>Library &amp; Infor. Sc.</td>
<td>7.88</td>
<td>3.64</td>
</tr>
<tr>
<td>Linguistics</td>
<td>6.96</td>
<td>3.00</td>
</tr>
<tr>
<td>Philosophy</td>
<td>7.14</td>
<td>3.43</td>
</tr>
<tr>
<td>Politics</td>
<td>6.62</td>
<td>2.28</td>
</tr>
<tr>
<td>Psychology</td>
<td>7.88</td>
<td>2.95</td>
</tr>
<tr>
<td>Sociology</td>
<td>8.24</td>
<td>3.53</td>
</tr>
</tbody>
</table>

Table 4.13: The mean and standard deviation (S.D.*) numbers of contents of Arabic non-scientific article titles in eleven disciplines.
4.3 A study of the information content of English titles

4.3.1 Objectives of the study

The objective of undertaking a new study of the information content of English titles is to compare the results with those of Arabic titles. One way to judge whether Arabic titles are informative is to compare the results with those of other titles which have been reported as informative. Section 4.1.1 reviewed many previous studies which revealed that English titles have been found to be sufficiently informative to be used for indexing; so the results of the Arabic titles will be compared with new results for English titles to find out whether a significant difference is occurring between them.

4.3.2 Method and samples

To undertake this new study of English titles, samples of 50 titles each were taken from 12 different subjects covering the period 1980 - 1988 in both scientific and non-scientific fields. Another four scientific subjects were borrowed from Buxton's studies (1979; 1987). These are medicine and chemistry (covering the years 1973 and 1984); physics (covering the year 1973); and engineering (covering the year 1974 - 1975). Physics and engineering were taken from Buxton's Ph. D. thesis (1979, p.45) whereas medicine and chemistry were taken from his 1987 paper. All the selected English periodicals have been agreed on with another research student as research journals comparable with the Arabic periodicals. These periodicals are shown in appendix 4.3.

It is relevant to mention that studying the information content of titles of English periodical articles was much easier to undertake than that of Arabic. This was because English titles had been investigated before by specialists
4.3.3 Results

The results of the study of the information content of titles of English periodical articles are shown in table 4.14 for titles taken from scientific disciplines, and in table 4.15 for titles selected from non-scientific disciplines. These tables provide the mean and standard deviation numbers for titles of the investigated subjects.

<table>
<thead>
<tr>
<th></th>
<th>All words</th>
<th>Substantive words</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.*</td>
</tr>
<tr>
<td>Agriculture</td>
<td>13.45</td>
<td>4.31</td>
</tr>
<tr>
<td>Chemistry</td>
<td>11.21</td>
<td>4.96</td>
</tr>
<tr>
<td>Engineering</td>
<td>11.00</td>
<td>4.70</td>
</tr>
<tr>
<td>Medicine</td>
<td>10.68</td>
<td>4.70</td>
</tr>
<tr>
<td>Physics</td>
<td>9.81</td>
<td>3.98</td>
</tr>
</tbody>
</table>

Table 4.14 : the mean and standard deviation (S.D.*) numbers of contents of English scientific article titles in five disciplines.
Table 4.15: The mean and standard deviation (S.D.*) numbers of contents of English non-scientific article titles in eleven disciplines.

<table>
<thead>
<tr>
<th>Subject</th>
<th>All words</th>
<th></th>
<th>Substantive words</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.*</td>
<td>Mean</td>
<td>S.D.*</td>
</tr>
<tr>
<td>Economics</td>
<td>9.90</td>
<td>3.54</td>
<td>5.20</td>
<td>1.08</td>
</tr>
<tr>
<td>Education</td>
<td>9.20</td>
<td>4.04</td>
<td>5.40</td>
<td>2.30</td>
</tr>
<tr>
<td>Geography</td>
<td>8.02</td>
<td>2.73</td>
<td>5.22</td>
<td>1.72</td>
</tr>
<tr>
<td>History</td>
<td>9.94</td>
<td>3.04</td>
<td>5.92</td>
<td>2.02</td>
</tr>
<tr>
<td>Law</td>
<td>7.20</td>
<td>3.80</td>
<td>3.50</td>
<td>1.79</td>
</tr>
<tr>
<td>Library&amp;Infor.Sc.</td>
<td>7.64</td>
<td>3.64</td>
<td>4.10</td>
<td>1.68</td>
</tr>
<tr>
<td>Linguistics</td>
<td>5.90</td>
<td>2.80</td>
<td>3.30</td>
<td>1.50</td>
</tr>
<tr>
<td>Philosophy</td>
<td>4.90</td>
<td>2.43</td>
<td>2.60</td>
<td>1.04</td>
</tr>
<tr>
<td>Politics</td>
<td>7.34</td>
<td>2.70</td>
<td>4.10</td>
<td>1.61</td>
</tr>
<tr>
<td>Psychology</td>
<td>8.92</td>
<td>3.80</td>
<td>4.70</td>
<td>1.71</td>
</tr>
<tr>
<td>Sociology</td>
<td>9.70</td>
<td>4.96</td>
<td>5.24</td>
<td>2.15</td>
</tr>
</tbody>
</table>
4.4 Comparison between the results of the two studies using the T test

The following tables 4.16 and 4.17 show the results which were obtained when the T test of significance was applied on titles of both Arabic and English articles. The test used was whether T was greater than 2.7, which has a probability of less than 1% of the difference arising by chance.

Most of the subjects showed no significant difference in the length or in the number of substantive words between both Arabic and English periodical article titles. The only exceptions were philosophy, which showed significant differences in both title length and substantive words, while agriculture, law, library and information science, and linguistics showed significant differences in the number of substantive words.

<table>
<thead>
<tr>
<th>Subject</th>
<th>All words</th>
<th>Substantive words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>0.22 A</td>
<td>4.33 A*</td>
</tr>
<tr>
<td>Chemistry</td>
<td>1.65 A</td>
<td>0.41 A</td>
</tr>
<tr>
<td>Engineering</td>
<td>0.00</td>
<td>0.40 E</td>
</tr>
<tr>
<td>Medicine</td>
<td>0.18 A</td>
<td>0.45 E</td>
</tr>
<tr>
<td>Physics</td>
<td>0.48 A</td>
<td>0.45 A</td>
</tr>
</tbody>
</table>

Table 4.16: values of T for differences between scientific titles of English and Arabic articles. A = Arabic are longer; E = English are longer; * = significant at 1% level.
<table>
<thead>
<tr>
<th>Subject</th>
<th>All words (T)</th>
<th>Substantive words (T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economics</td>
<td>0.70 E</td>
<td>2.12 A</td>
</tr>
<tr>
<td>Education</td>
<td>0.83 A</td>
<td>0.76 A</td>
</tr>
<tr>
<td>Geography</td>
<td>0.03 E</td>
<td>0.14 A</td>
</tr>
<tr>
<td>History</td>
<td>0.33 E</td>
<td>1.25 A</td>
</tr>
<tr>
<td>Law</td>
<td>2.42 A</td>
<td>5.22 A*</td>
</tr>
<tr>
<td>Library &amp; Infor. Sc.</td>
<td>0.32 A</td>
<td>2.87 A*</td>
</tr>
<tr>
<td>Linguistics</td>
<td>1.82 A</td>
<td>3.04 A*</td>
</tr>
<tr>
<td>Philosophy</td>
<td>3.76 A*</td>
<td>5.85 A*</td>
</tr>
<tr>
<td>Politics</td>
<td>1.44 E</td>
<td>0.79 A</td>
</tr>
<tr>
<td>Psychology</td>
<td>1.52 E</td>
<td>0.40 E</td>
</tr>
<tr>
<td>Sociology</td>
<td>1.69 E</td>
<td>0.53 A</td>
</tr>
</tbody>
</table>

Table 4.17: values of T for differences between non-scientific titles of English and Arabic articles. A = Arabic are longer; E = English are longer; * = significant at 1% level.
4.5 Discussion of the results for the two studies

Despite the fact that Arab authors have not been advised and directed yet as how to construct indicative titles for their articles, it was noticed that in all subjects which showed significant differences it was the Arabic titles which were longer. Obviously, this difference does not mean that Arabic titles are more representative than English titles but it does suggest that they are not less informative.

In this study, it was found that nearly all Arabic scientific subjects which were examined showed no significant difference in comparison with their English counterparts. The only exception was agriculture in Arabic. The significant difference between Arabic and English titles in the field of agriculture might be due to the fact that most of Arabic titles in agriculture contain country names, which in the Arab world are usually long and consist of more than one word. It seems that Arab authors prefer to write the full official name of any Arab country, although it can be represented by its widespread name which is normally constitutes one word. For example, the Kingdom of Saudi Arabia which is in Arabic (المملكة العربية السعودية) can be represented by one word (السعودية) al-Sa'udiyyah. Of the 50 samples of Arabic titles in agriculture which were investigated in this study, 28 contained country names and each is represented by at least two words. For instance the name of Qatar is given as the State of Qatar. And outside of this study, users may see the name Arab Republic of Syria used to mention Syria. Obviously, eliminating the word "State" and the words "Arab Republic" will not confuse the meaning.

In English subjects, the case is different because of the use of abbreviations where they are applicable, e.g. UK, USA, USSR, and the use of
one term for country names when it is possible. Of the 250 samples of Arabic scientific titles no single abbreviation was used.

Other scientific subjects like medicine, chemistry, physics and engineering appeared to have sufficient numbers of informative words in their titles, although some of these titles include terms in English script, e.g. (cyprinion mhalensis). Factors which might have contributed to this finding are the following:

a - The fact that many of the English scientific single terms are translated into two terms when they are used in Arabic titles, e.g. the term cervix (one word) is in Arabic عنق الرحم (two words), and the term thermodependence (one word) is in Arabic الإعتماد الحراري (two words). This will definitely increase the number of informative words in Arabic titles and it is an advantage when the process of retrieving by title words is concerned.

b - A second factor which seems possible is that Arab scientists who studied abroad had the chance to use English indexing and abstracting tools during their academic studies, and they might have been affected by the way Western authors write their titles since many of them wrote joint papers either with their supervisors or with their colleagues.

With regard to non-scientific subjects, significant differences are found in philosophy, library and information science, linguistics and law. A common practice in Arabic non-scientific titles that might have contributed to this difference is the use of compound terms and institution names which usually consist of more than three words and again the dearth of using abbreviations and acronyms. For instance the following title taken from the field of library and information science دور المنظمة العربية للتربية والثقافة والعلوم في استخدام الحاسب الإلكتروني في المكتبات ومراكز.
(المعلومات في الوطن العربي) contains an institution name of five keywords, whereas in the English translation of this title, which appeared with an abstract in the same journal, the institution's name was represented by an acronym (ALECSO). (Role of ALECSO in using computers in libraries and information centres in the Arab countries). Like most of the previously mentioned studies which were carried out on English titles, this study showed that article titles in scientific fields in Arabic are more informative than those in non-scientific fields.

4.6 Conclusion

Although indexing by each keyword in journal article titles was not properly considered or applied before, Arabic titles appear to be as informative as English titles in 16 scientific and non-scientific subjects. However, there are still two more fundamental issues to be investigated:

a - An informative title is not necessarily representative.

b - As stated by Buxton and Meadows (1977) "although sufficient information may be present in the titles it may not be in a form suitable for retrieval".

Further work on these two issues: representativeness and retrievability is undertaken in the following chapters. Two more practical problems in using Arabic titles are the need to strip inseparable prepositions from keywords, and the presence of some words in Roman and other non-Arabic languages script.
chapter five

Matching title keywords of Arabic articles and their corresponding subject headings in an interdisciplinary printed index

This chapter is devoted to comparing the Arabic article titles provided by authors and the indexer assigned subject headings to the articles. This chapter also presents a partial answer to the question of whether Arabic titles represent the contents of their accompanying articles.
5.1 Introduction

When title keywords indexes generated by computers were first being introduced, covering human indexers' deficiencies and offering considerable savings in time and cost, specialists' concerns centred on whether such printed indexes could perform as well as conventional indexes based on human assigned subject headings.

One of the early methods to test the adequacy of title keywords for indexing purposes was to match keywords found in titles with terms assigned to the corresponding articles in an available indexing or abstracting service. Bottle (1970) believed that "representative documents should be analysed for information content and each item of information checked against the document's title and judged whether or not it is deducible from the title". However, because "such a procedure would be very time consuming", he himself and others, before and after his studies, were content to follow an alternative procedure to examine the title's representation. That was to select a random sample of articles from an already analysed store, e.g. Index Medicus, and carry out a comparison between their title keywords and the corresponding indexer assigned subject headings.

5.2 Previous studies

A large number of studies have investigated the correlation between indexer-assigned indexing terms and title keywords. For the most part, these studies were concerned with the efficiency and effectiveness of article title keywords for indexing purposes. Some of these studies examined article titles in printed indexes and compared them with terms from their associated subject headings. Others were related to the introduction of computerized
information retrieval systems which enabled specialists to search profiles against titles and descriptors respectively, and then to compare the recall and precision obtained.

Montgomery and Swanson (1962), who are believed to be the first researchers to use the matching concept, studied entries in Index Medicus to determine the extent of match between keywords in the title and words identical to or synonymous with the corresponding subject headings. They used an approach based on the question, "to what extent can the human indexing operations that take place in an existing system be simulated by machine?". Their conclusion showed that if the computer was provided with the titles of the papers, MeSH (the Medical Subject Headings List), and a suitable synonym dictionary, it (the computer) could be programmed to determine which entry points should be assigned to each item. About 86% of such an assignment of entry points would be the same as the humanly selected subject headings by the indexers of the Index Medicus.

These results and conclusions contrast sharply with those of a parallel study carried out by O'Connor (1964a) of three other medical indexing systems. The results obtained ranged from 13% to 68% and led O'Connor to the general conclusion that "any proposal to replace a medical subject index produced by subject specialist with an automatic index based on 'thesaurus' processing of titles should be viewed with great caution".

These two studies and other similar works are comprehensively reviewed by Buxton (1979, p.26), Hann (1981, p.8), and Bachir (1987, p.119). Other recent matching studies that could be found do not in fact consider the presence of such matches to be indicators of the effectiveness of article title keywords. They are merely concerned with the overlap of title keywords and indexer assigned descriptors in CISTI's (the Canada Institute
for Scientific and Technical Information) OON database which contains monographs and technical reports in science, technology and medicine (Shepherd, 1987); and with the degree of match that exists between the controlled vocabulary of subject headings and the derived keywords within the titles of books (Frost, 1989).

5.3 Interdisciplinary studies

Most of the studies reviewed above and in other works are one discipline studies, and a fair number of them relate to the areas of biomedicine and chemistry. Two interdisciplinary studies which were reported in the literature are reviewed below because, despite the difference that exists in terms of language, size of samples, and degrees of matching, they are more or less similar to the new interdisciplinary matching study for Arabic articles. The process of conducting this study and its outcomes are reported in section 5.6.

Sedano (1964) carried out a study which compared title keywords and indexing terms for samples of 200 - 510 titles in six different indexing services. The comparisons were classed as:

a - The title of the article contained the index term or a word with the same root.

b - The title contained a synonym of the subject headings.

c - The title contained useful keywords but not those of the subject headings.

d - Non-descriptive. Sedano's outcomes are summarised in table 5.1.
Sedano's (1964) results of comparisons of title words and subject headings for a, b, c, d. See section 5.3.

Sedano suggested that the concept of a "technical field" is not limited to science, but that it might apply to any specialized field of knowledge. His study revealed that in the fields of science and engineering there is a 95% probability that between 79.3 and 92.5 of the titles contain sufficient keywords which clearly indicate the subject matter of the item. He pointed out that titles in these areas lend themselves well to KWIC indexing which is suitable and efficient for many highly specialized fields of knowledge.

Another interdisciplinary study was undertaken by Lane (1964) involving a larger number of indexing journals. He compared the suitability of article titles in various fields for KWIC indexing in nine indexing journals. Fifty titles were taken from each, then considered as acceptable titles if they contained a word which was identical or similar to the assigned terms under which they appeared in the indexing journals. If they did not, another similar check was done under the possible indexing terms. The acceptability of the titles in the nine indexing journals as reported by Lane are shown in table 5.2.

<table>
<thead>
<tr>
<th>Indexing journals</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engineering Index</td>
<td>83.1%</td>
<td>2.00%</td>
<td>12.8%</td>
<td>2.1%</td>
</tr>
<tr>
<td>Applied Sc.&amp;Tech.Index</td>
<td>88%</td>
<td>1.00%</td>
<td>7.5%</td>
<td>3.5%</td>
</tr>
<tr>
<td>ASM Review</td>
<td>86%</td>
<td>0.00</td>
<td>14%</td>
<td>0.00</td>
</tr>
<tr>
<td>Bulletin PAIS [public affairs]</td>
<td>55.3%</td>
<td>3.3%</td>
<td>37.3%</td>
<td>3.7%</td>
</tr>
<tr>
<td>Business Periodical Index</td>
<td>45%</td>
<td>7.5%</td>
<td>32%</td>
<td>15.5%</td>
</tr>
<tr>
<td>International Index [S.Sc.&amp;H]</td>
<td>32.2%</td>
<td>1.5%</td>
<td>44.2%</td>
<td>22.1%</td>
</tr>
</tbody>
</table>

Table 5.1 : Sedano's (1964) results of comparisons of title words and subject headings for a, b, c, d. See section 5.3.
Indexing journals | Percentage of "acceptable titles"
---|---
Applied Sc.&Tech.Index | 86%
Engineering Index | 82%
Bulletin of Public Affairs | 78%
Index to Legal Periodicals | 66%
Bibliography of Agriculture | 60%
Business Periodical Index | 58%
Education Index | 56%
International Index to Per. | 50%
Readers' Guide to Per. Lit. | 42%

Table 5.2: Lane's (1964) results of comparisons of title words and subject headings.

Lane’s general conclusion was that "in science and engineering the titles of the articles usually describe or imply the contents of the articles. In non-technical fields titles reveal the contents much less frequently; and in a general index such as Reader's Guide, titles are indicative less than half the time ".

5.4 Criticism of these studies

The studies surveyed above were criticized by several authors who demonstrated the fallacy of using manually assigned subject headings in conventional indexes and comparing them with mechanically derived terms from the titles of documents. O'Connor (1964b) observed that "an indexing
duplication investigation should not be called a test of mechanized indexing methods. For one can always ask " How good, really, is the human indexing being used as a standard? ". Hines and Harris (1970) pointed out that " machine or strict algorithmic selection of indexing words or phrases from titles or text is conceptually different from the essentially subjective methods of assigning indexing terms used by competent human indexers, which can not be described as rigorously ". They also mentioned that " there are too many variables and too many structural problems not taken into account by such procedures ". Another critical opinion was given by Feinberg (1972, p.69 ) who quoted Roberts (1963) to point out that since investigations of indexers have revealed that indexers rely to a great extent on titles for their subject analysis, evaluating titles by using associated humanly assigned subject headings loses its validity. She also pointed out that the decisions to include synonyms are subjective and questionable. In addition, she criticized the size of samples used in the investigations and concluded that "results based upon such sampling are open to question". In his thesis, Buxton (1979, p.37) mentioned another factor " which must have a considerable effect on particular experiments " that was " whether the indexing service has an existing authority list or whether it coins terms on an ad hoc basis ". He believes that in some indexing services, where free text is allowed, " indexers are likely to be influenced strongly by the title in their choices of terms ". Because all the studies reviewed in the literature considered the indexer assigned subject headings as a standard, Buxton (1979, p.36) referred to them as a relative ones and stated that " such index terms provide neither an ideal nor a constant standard ".

One more factor that also throws doubt on the reliability of human indexing, and which has already been mentioned by several studies, is the
inter-indexing and intra-indexing consistency. Leonard (1977) who surveyed most of such studies stated that since mid 1950, more than thirty of them have been carried out to measure inter-indexer consistency. These studies varied considerably in research methodology, in the number of documents selected, and in the number of indexers. They also varied considerably in the level of indexer consistency reported. Prior to Leonard's extensive study, Feinberg (1972, p.29) said that "all have concluded that manual indexing is inconsistent to varying degree whether performed by different indexers, or by the same indexer at different times ".

5.5 Criticism of the criticism

Despite the criticisms that have faced the studies which compared keywords selected from titles with human indexing terms, Harding (1982, p.4) still considers that the simplest method of evaluating computer generated indexes is "just to compare it to human - prepared indexing of the same document ". However, he keeps in mind that "this norm, while not necessarily "good " is the most objective method which can simply be defined ". Others questioned the indexing consistency as a factor that necessarily assures a good index. Saint Laurent (1967) raised many questions yet to be answered in regard to indexing consistency. Cooper (1969) noted that indexer consistency in itself is no guarantee of retrieval and that the indexing process can be consistently bad. Finally, Roberts' (1963) indication that indexers have a frequent tendency to rely on keywords appearing in titles should not have been considered by Feinberg (1972, p.69) and others, as a drawback of the method, especially if titles, which are parts of their corresponding items, are adequate and indicative.
5.6 New study for titles of Arabic periodical articles

Apparently, the literature of library and information science includes no previous studies that have compared keywords of titles of Arabic articles with their corresponding human-assigned subject headings. Such studies have not been carried out yet, either in a printed indexes environment or in computerized retrieval systems. Therefore, undertaking this new study would fill this gap in the general literature and open up discussions on titles of Arabic periodical articles and the effectiveness of their keywords in the retrieval process as well as the efficiency of assignment indexing for the Arabic literature. Unlike studies on English titles, this new study will not consider the occurrence of matches between title keywords of Arabic titles and the allocated subject headings as full indicators of the titles' representation of their accompanying articles. Another method is going to be undertaken as well in the following chapter to confirm the outcome of this matching process. In fact, apart from the criticisms that have already been made on the use of matching concept as a basis to investigate the representativeness of titles, Arab countries suffer from the shortage of well experienced and specialized indexers and from the weakness of the existing subject heading lists used in the indexing of Arabic materials.

5.7 Objectives of the study

The primary objective of this study is to find out the number of access points that users of current Arabic printed indexes would lose or gain in retrieving specific documents in specific fields through the keywords in their titles rather than the indexer assigned subject headings. The investigations to fulfill this objective would also reveal the average number of subject headings allocated to each article, in various subjects, by the human indexers
as well as the average number of access point keywords found in each article title in the same subjects.

5.8 Method and samples

Three issues are going to be dealt with under this section. First, the number of sample articles used and the way they were gathered; second, the matching categories which will be used in the comparison process; and third, the definitions of each category with examples from the real samples.

5.8.1 Sample determination

To accomplish this study, 234 article titles with their corresponding humanly assigned subject headings in 10 different fields were obtained from the institution that produces the leading conventional printed index in the Arab world, al-Fihrist (الفهرست). See section 2.3.2.B.2.

Initially, indexers in al-Fihrist were asked to supply at least 25 complete indexing sheets in each of the 16 scientific and non-scientific subjects examined in chapter four, but unfortunately their policy is not to keep indexing sheets although al-Fihrist is not stored on any electronic device or in computerized retrieval systems. They could only supply this study with 234 indexing sheets covering 10 fields for the period between 1986 and 1988. Table 5.3 shows the number of indexing sheets taken from each of the 10 fields with their allocated subject headings which appeared on the supplied indexing sheets. To see the journal sources from which these samples were originally taken, see appendix 5.1.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Indexing sheets</th>
<th>Subject headings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>17</td>
<td>50</td>
</tr>
<tr>
<td>Economics</td>
<td>25</td>
<td>79</td>
</tr>
<tr>
<td>Education</td>
<td>30</td>
<td>106</td>
</tr>
<tr>
<td>Geography</td>
<td>18</td>
<td>53</td>
</tr>
<tr>
<td>History</td>
<td>26</td>
<td>96</td>
</tr>
<tr>
<td>Law</td>
<td>25</td>
<td>81</td>
</tr>
<tr>
<td>Philosophy</td>
<td>23</td>
<td>82</td>
</tr>
<tr>
<td>Politics</td>
<td>28</td>
<td>105</td>
</tr>
<tr>
<td>Psychology</td>
<td>25</td>
<td>85</td>
</tr>
<tr>
<td>Sociology</td>
<td>17</td>
<td>52</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>234</strong></td>
<td><strong>789</strong></td>
</tr>
</tbody>
</table>

Table 5.3: number of sample articles and their allocated subject headings in 10 subjects taken form *al-Fihrist*.

It can be observed in the above table that the number of subjects involved in this study is not fairly divided between scientific and non-scientific fields, and the study seems to be concentrating on author's titles and indexer's subject headings in non-scientific subjects, with agriculture being an exception.

In fact this limitation in the number of scientific fields investigated in this study arises from three main reasons:

a - No index is currently available that covers the contents of Arabic scientific periodicals.

b - *Al-Fihrist* index is a non-scientific interdisciplinary index with the
exception of covering articles on agriculture.

c - The Arabic scientific articles are not covered or stored in any printed or computerized retrieval systems in the Arabic language.

5.8.2 Categories of analysis

In this study, several categories of similarity between keywords in Arabic titles and humanly assigned subject headings were identified, and these categories correspond to the levels of match addressed by the following questions:

a - To what extent are title keywords exactly the same as their corresponding indexer-assigned subject headings in al-Fihrist?

b - To what extent are title keywords the same as the first word which provides access point in their corresponding subject headings?

c - To what extent do keywords in Arabic article titles match parts of the access points in their corresponding subject headings?

d - To what extent are keywords in Arabic article titles synonymous with the access points in their corresponding subject headings?

e - To what extent are keywords in Arabic article titles not indicated in their corresponding subject headings?

Although the keywords have been divided into five classes, a, b, c, d, and e, classes a, b, c, and d, together represent a single grouping of those articles that are theoretically retrievable from a keyword index if all possible inflexional variants are considered and all possible synonyms are used. Class e in itself represents the opposite case where there is no possibility of access at all.
5.8.3 Definitions of categories with Arabic examples

Definitions of what constitutes a match have varied from one study to another. Most studies have defined different degrees of matching and assigned matching categories based on these different degrees. Generally, three types of categories are used. These are defined as: exact match, partial match, and no match. However, in this study six categories were used.

The following are the six matching categories adopted for this study, with their definitions and some examples from the Arabic literature alongside the English translations:

5.8.3.A Exact match: entire heading

The category of exact matchings shows how similar title keywords are to Arabic subject headings vocabulary. In an exact match the entire subject heading was represented by the keywords in titles. The complete subject heading could consist of a main term with or without subdivision. For example,

Title: مهندسة الجغرافيا الاقتصادية
Subject heading: الجغرافيا الاقتصادية - المنهجية

English translation:
Title: Methodology of economic geography
Subject heading: Economic Geography - Methodology

5.8.3.B Exact match: access point only

In this category, at least one of the title keywords matches exactly the access point of a subdivided subject heading. For example,

Title: تطوير طرائق التدريس في اللغات الحية
Subject heading: Languages - study and teaching

English translation:
Title: Development of teaching methods for live languages.
Subject heading: Languages - study and teaching

5.8.3.C Stem match: access point

In a stem match, only part of the subject heading's access point matches at least one of the title keywords. Singular and plural variants, introductory prefixes and suffixes, were all considered and attended to during the matching process. An example,

Title: دراسة مستوى الرضى والروح المعنوية لموظفين جامعة موتاه
Subject heading: الموظفون - الأردن - الاختبارات والمقاييس

English translation:
Title: A study of satisfaction and morale among the employees of Mu'tah University
Subject heading: Employees - Jordan - test and measures

Here, the English translation of the Arabic title does not show the grammatical difference between the appearance of the keywords لموظفي (li-muwazzafi) in the title, and the access point الموظفون (al-muwazzafun) in the subject heading. In fact, due to their respective positions in the title and in the subject headings, their endings have become different from each other.

5.8.3.D Synonym match: access point

In this category, access points of subject headings which synonymously matched one of the title keywords were counted. Synonym here includes words which are different in terms of forms and spellings but equal in terms of meanings. For example,
Title: The required qualifications in secondary school teachers
Subject heading: Instructors - Iraq - tests and measures

5.8.3.E No match

This category included the number of subject headings which had no match at all levels with the title keywords. For example,

Title: Methodology of economic geography
Subject heading: Marxism

To these five classes, another sixth was added to investigate matches between keywords in titles and words in subdivisions. This class has no role in the matching study between title keywords of Arabic periodical articles and the access points of their corresponding subject headings. However, it can be of vital importance for the purpose of revealing the degree of similarity between keywords in titles and keywords in subject headings. This category is ranked F and defined as the following:

5.8.3.F Subdivisions match

A match on subdivisions would in fact be very useful towards the study of titles' representation of their corresponding articles, if the subdivisions are topical or geographical. For example,
There are cases where, in addition to the keyword matching the access point of a subject heading, there is also a keyword matching the subdivision. For example,

Title: The relation between local currency surplus and balance of payments deficit in non-oil producing developing countries: 1968 - 1983

Subject heading: Developing countries - balance of payments - principles and theories

No special categories were designed for these matches, but occurrences of such cases were mentioned alongside their particular categories, instead.

Since indexing terms in al-Fihrist are complex, they were separated into their component parts of keywords and those parts used to carry out the process of comparison with keywords in titles of Arabic articles. Cross-references as well as punctuation and voweling marks were disregarded in this study, and direct order of keywords was dropped from the conditions of forming an entire match, except in cases where the meaning would be affected. In this study, assigned subject headings were ranked according to their various level of match with the titles of articles. For instance, if a title contained keywords matching an entire subject heading as
well as an access point of another one, the title was counted as an entire exact match for the first and as an exact match of access point for the second. For example, the title "Wages in Lebanon" is an entire exact match for the subject heading "Wages - Lebanon", and an exact match for access point in the subject heading "Wages - Arab world - statistics".

So, each subject heading was put in one matching category and in cases where more than one occurrence exists, it was ranked according to its highest level of match. Conventions of what to consider a keyword and a stopword are the same of those which were adopted previously in chapter four. Finally, a subject heading was defined as every term or set of terms with or without subdivisions assigned by a human indexer regardless of whether it was taken from a subject heading list, thesaurus, or it was assigned on an ad hoc basis. As for al-Fihrist, indexers depend on a subject heading list which was developed for their own use. For this study only subject headings which appeared on the back of al-Fihrist indexing sheets were counted. In fact, instructions were given by one of the indexers in al-Fihrist that the back of their indexing sheet is allocated for subject headings, whereas the front is for other details such as title, author and bibliographic details of the articles.

5.9 Results and analysis

Table 5.4 lists the number and percentage of access points in subject headings which matched keywords in titles, divided into four categories in 10 different subjects. These categories are referred to as: A - for entire exact matches between title keywords and their corresponding subject heading, and A% for the percentage of subject headings in category A; B - for access point exact matches between title keywords and their corresponding subject heading, and B% for the percentage of subject headings in category B; C -
for stem matches of access point between title keywords and their corresponding subject heading, and C% for the percentage of subject headings in category C; and D - for synonym matches of access point between title keywords and their corresponding subject heading, and D% for the percentage of subject headings in category D. The other two categories which are also listed in table 5.4 are E - for the number of subject headings which had no matches at all with title keywords and E% for the percentage of the non-matching cases; and finally F - for the number of matches which occurred between title keywords and the subdivision of their corresponding subject heading, and F% for the percentage of subject headings' subdivisions in category F.
<table>
<thead>
<tr>
<th>Subject</th>
<th>A</th>
<th>A%</th>
<th>B</th>
<th>B%</th>
<th>C</th>
<th>C%</th>
<th>D</th>
<th>D%</th>
<th>E</th>
<th>E%</th>
<th>F</th>
<th>F%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>16</td>
<td>32%</td>
<td>02</td>
<td>04%</td>
<td>07</td>
<td>14%</td>
<td>02</td>
<td>04%</td>
<td>03</td>
<td>06%</td>
<td>20</td>
<td>40%</td>
</tr>
<tr>
<td>Economics</td>
<td>15</td>
<td>18.9%</td>
<td>14</td>
<td>17.7%</td>
<td>09</td>
<td>11.3%</td>
<td>01</td>
<td>01.2%</td>
<td>19</td>
<td>24.1%</td>
<td>21</td>
<td>26.6%</td>
</tr>
<tr>
<td>Education</td>
<td>11</td>
<td>10.4%</td>
<td>20</td>
<td>18.9%</td>
<td>06</td>
<td>05.6%</td>
<td>07</td>
<td>06.6%</td>
<td>26</td>
<td>24.6%</td>
<td>36</td>
<td>34%</td>
</tr>
<tr>
<td>Geography</td>
<td>14</td>
<td>26.4%</td>
<td>05</td>
<td>09.4%</td>
<td>13</td>
<td>24.5%</td>
<td>-</td>
<td>0</td>
<td>11</td>
<td>20.8%</td>
<td>10</td>
<td>18.8%</td>
</tr>
<tr>
<td>History</td>
<td>12</td>
<td>12.5%</td>
<td>17</td>
<td>17.7%</td>
<td>08</td>
<td>08.3%</td>
<td>04</td>
<td>04.2%</td>
<td>33</td>
<td>34.4%</td>
<td>22</td>
<td>22.9%</td>
</tr>
<tr>
<td>Law</td>
<td>13</td>
<td>16%</td>
<td>17</td>
<td>21%</td>
<td>13</td>
<td>16%</td>
<td>02</td>
<td>02.5%</td>
<td>19</td>
<td>23.5%</td>
<td>17</td>
<td>21%</td>
</tr>
<tr>
<td>Philosophy</td>
<td>26</td>
<td>31.7%</td>
<td>09</td>
<td>11%</td>
<td>04</td>
<td>04.8%</td>
<td>01</td>
<td>01.2%</td>
<td>35</td>
<td>42.7%</td>
<td>07</td>
<td>08.5%</td>
</tr>
<tr>
<td>Politics</td>
<td>08</td>
<td>7.6%</td>
<td>27</td>
<td>25.7%</td>
<td>11</td>
<td>10.5%</td>
<td>01</td>
<td>0.9%</td>
<td>30</td>
<td>28.6%</td>
<td>28</td>
<td>26.7%</td>
</tr>
<tr>
<td>Psychology</td>
<td>10</td>
<td>11.7%</td>
<td>15</td>
<td>17.6%</td>
<td>04</td>
<td>04.7%</td>
<td>05</td>
<td>05.8%</td>
<td>28</td>
<td>32.9%</td>
<td>23</td>
<td>27%</td>
</tr>
<tr>
<td>Sociology</td>
<td>10</td>
<td>19.2%</td>
<td>02</td>
<td>03.8%</td>
<td>05</td>
<td>09.6%</td>
<td>03</td>
<td>05.7%</td>
<td>25</td>
<td>48.1%</td>
<td>07</td>
<td>13.5%</td>
</tr>
</tbody>
</table>

Table 5.4: number and percentage of access points and subdivisions for subject headings which matched keywords in titles of Arabic articles.
The overall or total number of access points in subject headings that matched exactly or partially keywords in titles of Arabic periodical articles, is shown in table 5.5 represented by the letter G for the number of matches and G% for the percentage of subject headings in category G. Table 5.6 on the other hand lists the total number of similarities between words in subject headings (including inflexional variations and synonyms) and keywords found in the investigated titles. It constitutes 6 columns: G and G% are as indicated before; F and F% are respectively designated for the number of matches on subdivisions and the percentage of subdivisions in category F. Whereas H and H% are devoted to show the total number of similarities between title keywords and their corresponding subject headings in the 10 various subjects, and the percentage of this similarity. Table 5.7 shows the average of subject headings assigned to each article in the 10 subjects as well as the average number of keywords found in every article title in the same 10 subjects. The purpose of this table is to reveal the depth of indexing in *al-Fihrist* through the number of subject headings; and whether titles of Arabic articles provide more access points to users than the leading conventional printed index in the Arab World.
<table>
<thead>
<tr>
<th>Subject</th>
<th>G</th>
<th>G%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>27</td>
<td>54%</td>
</tr>
<tr>
<td>Economics</td>
<td>39</td>
<td>49%</td>
</tr>
<tr>
<td>Education</td>
<td>44</td>
<td>41.5%</td>
</tr>
<tr>
<td>Geography</td>
<td>32</td>
<td>60.3%</td>
</tr>
<tr>
<td>History</td>
<td>41</td>
<td>41.6%</td>
</tr>
<tr>
<td>Law</td>
<td>45</td>
<td>55.5%</td>
</tr>
<tr>
<td>Philosophy</td>
<td>40</td>
<td>47.5%</td>
</tr>
<tr>
<td>Politics</td>
<td>47</td>
<td>44.7%</td>
</tr>
<tr>
<td>Psychology</td>
<td>34</td>
<td>39.8%</td>
</tr>
<tr>
<td>Sociology</td>
<td>20</td>
<td>38.4%</td>
</tr>
</tbody>
</table>

Table 5.5: overall number of access points which matched keywords in Arabic titles and their percentage.
<table>
<thead>
<tr>
<th>Subject</th>
<th>G</th>
<th>G%</th>
<th>F</th>
<th>F%</th>
<th>H</th>
<th>H%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>27</td>
<td>54%</td>
<td>20</td>
<td>40%</td>
<td>47</td>
<td>94%</td>
</tr>
<tr>
<td>Economics</td>
<td>39</td>
<td>49.1%</td>
<td>21</td>
<td>26.6%</td>
<td>60</td>
<td>75.9%</td>
</tr>
<tr>
<td>Education</td>
<td>44</td>
<td>41.5%</td>
<td>36</td>
<td>34%</td>
<td>80</td>
<td>75.5%</td>
</tr>
<tr>
<td>Geography</td>
<td>32</td>
<td>60.3%</td>
<td>10</td>
<td>18.8%</td>
<td>42</td>
<td>79.2%</td>
</tr>
<tr>
<td>History</td>
<td>41</td>
<td>42.7%</td>
<td>22</td>
<td>22.9%</td>
<td>63</td>
<td>65.6%</td>
</tr>
<tr>
<td>Law</td>
<td>45</td>
<td>55.5%</td>
<td>17</td>
<td>21%</td>
<td>62</td>
<td>76.5%</td>
</tr>
<tr>
<td>Philosophy</td>
<td>40</td>
<td>48.7%</td>
<td>07</td>
<td>08.5%</td>
<td>47</td>
<td>57.3%</td>
</tr>
<tr>
<td>Politics</td>
<td>47</td>
<td>44.7%</td>
<td>28</td>
<td>26.7%</td>
<td>75</td>
<td>71.4%</td>
</tr>
<tr>
<td>Psychology</td>
<td>34</td>
<td>39.8%</td>
<td>23</td>
<td>27%</td>
<td>57</td>
<td>67.1%</td>
</tr>
<tr>
<td>Sociology</td>
<td>20</td>
<td>38.4%</td>
<td>07</td>
<td>13.5%</td>
<td>27</td>
<td>51.9%</td>
</tr>
</tbody>
</table>

Table 5.6: Number and percentage of similar words between title keywords of Arabic articles and their corresponding subject headings in 10 subjects.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Keywords per title</th>
<th>Subject headings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>6.82 2.04</td>
<td>2.94 1.30</td>
</tr>
<tr>
<td>Economics</td>
<td>5.08 2.13</td>
<td>3.16 1.00</td>
</tr>
<tr>
<td>Education</td>
<td>5.80 2.00</td>
<td>3.53 1.43</td>
</tr>
<tr>
<td>Geography</td>
<td>5.50 2.17</td>
<td>2.94 1.27</td>
</tr>
<tr>
<td>History</td>
<td>5.77 1.89</td>
<td>3.69 1.03</td>
</tr>
<tr>
<td>Law</td>
<td>5.20 2.26</td>
<td>3.24 1.07</td>
</tr>
<tr>
<td>Philosophy</td>
<td>4.39 1.55</td>
<td>3.56 1.38</td>
</tr>
<tr>
<td>Politics</td>
<td>4.68 1.69</td>
<td>3.75 0.99</td>
</tr>
<tr>
<td>Psychology</td>
<td>5.72 2.90</td>
<td>3.40 0.63</td>
</tr>
<tr>
<td>Sociology</td>
<td>4.29 1.60</td>
<td>3.06 1.21</td>
</tr>
</tbody>
</table>

Table 5.7: average number of keywords per title and subject headings per article in 10 fields. * = standard deviation
5.10 Discussion

Under this section, the results obtained from the adopted categories are going to be discussed as separate categories and then as combined ones.

5.10.1 Single categories

5.10.1.A Exact match : entire heading

17.1% of the total number of subject headings found on al-Fihrist indexing sheets had entire exact matches with their corresponding article titles. The range of matches within the various fields was from 7.6% to 32%. The highest percentage of entire exact match category was found in the only scientific field investigated in this study: that was agriculture. The other nine subjects forming the non-scientific group had a combined percentage of 16.1% for entire exact matches. In this group the highest degree was that of philosophy which had 31.7%. The second highest percentage of entire exact matches was found in geography which had 26.4% followed by 19.2% for sociology. At the end of the line was politics which had 7.6% of the entire subject heading exact matches.

5.10.1.B Exact match : access point only

The percentage of matches in this category came lower than the category of exact matches for an entire subject heading. Initially, it was anticipated the percentage of matches for access points only would be larger than in the category of exact matches for complete subject heading because indexing concepts in al-Fihrist are complex and often have two subdivisions and the subdivisions did not have to match in this case. On the contrary, 16.2% of the subject headings matched title keywords of Arabic periodical
articles. This percentage, is almost identical to the 17.1% match on entire subject headings but is still a lower one.

In contrast with the category of exact matches for the full subject headings, where agriculture enjoyed the highest rate of match, the leader in this category of access point matches was the field of politics which had 25.7%. Next in line was the field of law which had 21% followed by the fields of education which had 18.9% and economics and history which had the same rate of matches 17.7%. The lowest rate of match in this category was for sociology which got 3.8%.

Some of the subjects which had the lowest degree of exact match on entire subject headings had the highest rate of exact match on access point of subject headings. All of the highest percentages of match in this category were found in the group of non-scientific disciplines. Politics had a match of 25.7% and education had 18.9% whereas in the previous category they represented the lowest rate of match and respectively had 7.6% and 10.4%. Agriculture which enjoyed the highest rate of exact match for entire headings was in this category among the subjects which had a low rate of match. It had a 4% match rate, whereas for the entire heading it had 32%.

Since the field of politics had a relatively high percentage in this category, it was examined whether the occurrences of names or places had been a factor. In fact, this proved to be the case, especially in an index like al-Fihrist where indexers prefer to use country names as main headings rather than subdivisions in most of their assignments for indexing concepts. Within 54 subject headings in politics, 38.9% of the matches on access points were names of places. This also showed that politics and education had more complex subject headings, in other words had more subdivisions in their headings.
5.10.1.C **Stem match : access point**

Overall, the category C matches formed 10.1%. Within disciplines, the highest rate of match in this category was for geography which had 24.5% of the rate of match. Second was law which had 16%, followed by agriculture 14% and economics 11.3%. Of the 24.5% matches in geography 43.5% occurred due to suffix variations, e.g. ( البيئة/ البيئة = environment / environmental), and 25% due to infixes changes. The lowest rate was 4.7% in psychology. Apart from the definite article in Arabic 'al ( آل ) which is written prefixed to the keyword it defines and which was not counted in this category, no prefixes variations were detected.

5.10.1.D **Synonym match : access point**

The percentage of matches in this category was very low. Only 3.3% of the samples had their access points in subject headings synonymously matching title keywords. The field of education enjoyed the highest rate of synonym match. It had 6.6% of the match in this category. Next was psychology reaching 5.8% and sociology which got 5.7% of the matches.

5.10.1.E **No match**

For 29% of the sample, there were no words from the subject headings which matched exactly or partially any of the keywords in titles. The highest rate of non-matches was 48.1% in sociology. Second highest was the field of philosophy which had 42.7% followed by history which had 34.4%. The lowest were 6% in agriculture and geography which had 20.8%.

In this category, agriculture which had the highest mean number of keywords in its titles (see table 5.7) had the lowest percentage of the
non-matches, whereas sociology and philosophy which respectively had the highest percentage of non-matches, had the lowest mean number of keywords in their titles. So, irrespective of the number of assigned indexing concepts, the higher the number of keywords in titles is, the more are the occurrences of matches with words in subject headings of their corresponding articles. This is not surprising, as expected.

5.10.1.F Subdivisions match

Throughout the total sampling, subject heading subdivisions matches occurred in 24.4%. For almost all fields, the percentages in this category were relatively high. This was probably because keyword matching had been considered in its broad term with no restrictions, for instance whether the match was exact or partial. Scientific fields represented by agriculture had the highest rate of match, that is 40% followed by the social science fields education and psychology which respectively had 34% and 27%. This category was only added to the previous ones to enable us to have the percentage of total similarities between keywords in titles and words in subject headings.

5.10.2 Combined categories

A review of all the matching categories used in this study indicates that 46.7% of the total sampling of subject headings had their access points matched with title keywords and that 71% of the subject headings contained at least one keyword in each article title, with the synonyms and inflexional variations considered (see table 5.9). If the category of synonym match D which has been one of the main controversial issues in the studies concerned with comparison of title keywords of articles and their corresponding
indexing concepts, was dropped, the percentage of matching words would decrease to 67.6% and if the category of subdivisions match F was excluded, the outcome would be that 43.4% of al-Fihrist subject headings had their access points exactly or partially similar to title keywords from 10 different scientific and non-scientific fields. Whereas table 5.8 shows the percentage of similar words H% which occurred in each of the subjects examined, table 5.9 lists the number and percentage of subject headings in each of the six matching categories used in this study. Finally, table 5.10 provides the number and percentage of matches according to two discipline groups; scientific represented by agriculture and non-scientific represented by the remaining 9 fields.
<table>
<thead>
<tr>
<th>Rank</th>
<th>Subject</th>
<th>H%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Agriculture</td>
<td>94%</td>
</tr>
<tr>
<td>2</td>
<td>Geography</td>
<td>79.2%</td>
</tr>
<tr>
<td>3</td>
<td>Law</td>
<td>76.5%</td>
</tr>
<tr>
<td>4</td>
<td>Economics</td>
<td>75.9%</td>
</tr>
<tr>
<td>5</td>
<td>Education</td>
<td>75.5%</td>
</tr>
<tr>
<td>6</td>
<td>Politics</td>
<td>71.4%</td>
</tr>
<tr>
<td>7</td>
<td>Psychology</td>
<td>67.1%</td>
</tr>
<tr>
<td>8</td>
<td>History</td>
<td>65.6%</td>
</tr>
<tr>
<td>9</td>
<td>Philosophy</td>
<td>57.3%</td>
</tr>
<tr>
<td>10</td>
<td>Sociology</td>
<td>51.9%</td>
</tr>
</tbody>
</table>

Table 5.8: percentage of similarities between title keywords and words in subject headings in 10 fields.
<table>
<thead>
<tr>
<th>Categ.</th>
<th>Type of match</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Exact match</td>
<td>135</td>
<td>17.1%</td>
</tr>
<tr>
<td>B</td>
<td>Exact match of acc.point</td>
<td>128</td>
<td>16.2%</td>
</tr>
<tr>
<td>C</td>
<td>Stem match of acc.point</td>
<td>80</td>
<td>10.1%</td>
</tr>
<tr>
<td>D</td>
<td>Synon.match of acc.point</td>
<td>26</td>
<td>03.3%</td>
</tr>
<tr>
<td>E</td>
<td>no match</td>
<td>229</td>
<td>29%</td>
</tr>
<tr>
<td>F</td>
<td>Subdivisions match</td>
<td>191</td>
<td>24.2%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>789</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

Table 5.9: number and percentage of subject headings according to categories of matches.
<table>
<thead>
<tr>
<th>Groups (categories)</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scientific (A-D and F)</td>
<td>47</td>
<td>94%</td>
</tr>
<tr>
<td>Non-scientific (A-D and F)</td>
<td>513</td>
<td>69.4%</td>
</tr>
<tr>
<td>Scientific (A-D only)</td>
<td>27</td>
<td>54%</td>
</tr>
<tr>
<td>Non-scientific (A-D only)</td>
<td>342</td>
<td>46.3%</td>
</tr>
</tbody>
</table>

Table 5.10: number and percentage of matches according to their discipline groups.

5.11 Conclusion

No previous studies have compared title keywords of Arabic periodical articles with their corresponding subject headings. However, some have approached the issue of subject cataloguing and subject heading lists used in the Arab countries.

Irrespective of the limitations which surrounded some of its aspects, this study paves the way for some Arab researchers to build on the work of others. As far as limitations are concerned, some of them were beyond the capacity of this study. One field had to be representative of scientific disciplines, simply because other fields were inaccessible and no indexing tool was available. The variation in the size of samples from one subject to another was imposed, because al-Fihrist index was the only available source of samples and subject headings. In this study, the matching categories were defined to apply to Arabic printed indexes, however one of them, subdivisions match, is relevant only to online systems with keyword searching, which is beyond the scope of this research.

The primary objective of this study had been to find out the degree of
similarity between title keywords of Arabic periodical articles and their corresponding subject headings' access points. The secondary objective was to reveal the average number of access points (keywords) in titles of Arabic periodical articles compared with the average number of their subject headings. Converting the obtained figures of access point match categories (excluding synonym matches category) to percentages revealed that 43.4% of the 789 subject headings humanly assigned to 234 articles had their access points similar to keywords in titles of Arabic periodical articles. Including synonym matches would augment the percentage to nearly half the subject headings (46.7).

Keeping in mind the weakness in the al-Fihrist authority list mentioned earlier in chapter two, and the fact that a great many of the access point matches were names of either geographical places or individuals (agriculture, politics), it is believed this percentage is encouraging towards depending on Arabic title keywords for the production of printed indexing services to disseminate current information to Arab users.

Other findings showed that al-Fihrist is not really an indepth indexing tool. The lowest mean number of title keywords, represented by sociology, is higher than the highest average number of subject headings found in politics. In all selected subjects, access points (keywords) found in titles were more than their humanly allocated subject headings. The overall average number of access points provided by the article titles in the ten fields was 5.33. Whereas their average of subject headings was 3.33.

Provided that the matches in the subdivisions match category are representative of their corresponding articles, but their prevention from being main terms was due to construction rules of subject headings or authority list used for al-Fihrist, it is believed that the results obtained in most of the fields
investigated are encouraging and promising towards the possibility of relying on title keywords for the purpose of producing printed indexes for Arabic literature, especially in agriculture, law, geography, economics, education and politics. However, since these results are relative (as title keywords were evaluated against indexer assigned terms), further investigation of title's representativeness and retrieval evaluation from an Arabic title-derivative printed index are needed. These two issues will be treated respectively in chapters six and eight.
Chapter six

The occurrence of title keywords in their article's topic sentences

This chapter is devoted to studying the representation of Arabic titles of the contents of their accompanying articles. For this, it uses a new method which relies on the principle of matching between the keywords of Arabic article titles and the keywords found in the topic sentences. Before starting the matching process, this chapter defines the term topic sentence and reviews the previous use of this concept in the field of librarianship.
6.1 Introduction

The previous chapter showed that the process of comparing keywords of article titles to their corresponding assigned subject headings has faced criticisms by some specialists in the field. One of the primary criticisms was that human indexers are by no means an ideal standard to investigate a title's representation. Consequently, the encouraging results that were obtained after the Arabic title keywords and their subject headings were matched, are affected by these criticisms and the limitation of the process.

One way to find out whether these criticisms are really affecting the results and distorting the conclusion is to carry out a different technique that eliminates such limitations. The new outcome would either confirm the previous results, so that the criticisms were trivial, or contradict them and prove the distortion. To overcome such criticisms, this chapter will be carrying out a different method using a totally different technique that eliminates the human indexer factor. This was a main element in almost every criticism put by those who disagreed with the process of matching title keywords with their corresponding conventional subject headings as a sole indication for the title's representation.

6.2 Objectives

Ideally, the representation of a text is the full text itself, and the study of title keywords representation of their accompanying article should involve each word being checked against the title with which it appears. Because such a process is time-consuming and currently not applicable automatically for Arabic texts, an alternative method will be adopted. This method involves the detection of title keywords in the topic sentences found in the
introductory or objectives parts of an article. The aim of this chapter is to look for each keyword in the titles of the selected Arabic articles and then to match them with the keywords of the topic sentences humanly detected in the introduction or objectives parts of those articles. It is obvious that the performance of this new technique overcomes the noted criticisms, since it is putting aside the human indexers' effect which applied in the previous method.

6.3 Usage of topic sentences

Under this section, firstly the circumstances in which the concept topic sentence will be used in this chapter are defined. The assumption behind its usage as an information rich segment which is considered very useful to investigate the relevance of Arabic titles to their corresponding articles, is justified. Secondly the previous use of topic sentences in the field of indexing will be reviewed and the way they were defined and used by specialists in producing automatic abstracts and/or extracts.

6.3.1 The use of topic sentences in this study

For the purpose of this study a topic sentence was defined as any string of words starting with specific phrases such as " in this study "، " this article "، " the following paper "； terminated by a full stop, question mark, or semicolon, and located in the introduction or objectives parts of an article. A list of all such Arabic phrases found at the beginning of the studied Arabic topic sentences is displayed in appendix 6.1. It is believed that when an author of an article includes in a sentence such a phrase, it is certain that he is turning from his review of the background or the previous studies to what he, himself, is going
to attempt and what his article is all about. Such sentences usually state what exactly is the subject of the study.

The assumption behind the dependence on topic sentences as a means to examine the representation by Arabic titles of their accompanying articles relies on the fact that topic sentences are used by authors to indicate the subject matter of their articles. Therefore, they are assumed to be sentences that most represent the content of their documents. In fact, topic sentences are not and cannot be the ideal representation of the whole accompanying texts, but they certainly can be considered as the best among other parts or entities used to represent the content of articles. In the literature of indexing, there is a basic assumption that human indexers are often able to state what an article is about by formulating what is defined as subject headings or index main entries. If human indexers are sometimes successful in representing the documents which they read by means of words or sets of words taken from prespecified lists, then there is no doubt that authors of texts are more successful in formulating sentences which explore the aboutness of their articles. These sentences are in fact some kind of semantic condensation prepared by those who wrote their amplifications (the original full text). So all variables which can affect the understanding of an article's text by an external human indexer will disappear in this process.

Another important factor which might be used to justify the dependence on topic sentences is the fact that authors are not supposed to use such sentences to deceive the reader's thinking, and the provision of unindicative provocative topic sentences will not serve the purpose of attracting readers. Unlike titles which sometimes are used for this purpose, topic sentences do not appear on periodical covers nor in lists of contents. On the contrary, a reader might have to read several paragraphs before getting to them, and this
is possible only in cases where they are used. Topic sentences are not a must in the structure of the text of an article or paper like the titles for instance.

One more observation which justifies the dependence on topic sentences as a means to investigate the relevance of Arabic articles was the fact that many of these sentences found in the selected articles were circled or highlighted by the indexers of al-Fihrist. It is mentioned later in this chapter that some of the samples studied were ordered from al-Fihrist.

Buxton and Meadows (1978) stated that the topic sentences will normally be "an amplification of the title, and contain several of the same keywords". Every item of information, whether it is a document, an article, or a paper, has to have a topic or a subject as well as a title to meet the least conditions of publishing. In this chapter, the idea is if titles of Arabic articles are representative and topic sentences are used to indicate the subject matter of those articles, then keywords of titles and topic sentences should be identical. This identification is reflected by the occurrence of either the same or synonymous keywords in both texts.

6.3.2 Previous usage of topic sentences

In the early days of computerized indexing and abstracting, specialists in the field concentrated on the problem of finding and defining sentences which could be extracted from a text of a document to convey a good idea of its content in order to produce automatic abstracts. The first experiment on automatic abstracting was conducted by Luhn (1958) and was called the keyword method. This and all other subsequent studies carried out up to about 1970 aimed to produce extracts by using key sentences automatically selected to represent the subject matter of the document and then printed in their order of occurrence to get an automatic extract. Lancaster (1991, p.222)
mentioned that various methods for identifying "information rich" segments or parts of document texts were suggested, and that computer programs would search for such parts as "prepositional phrases, text following" clue words like conclusions and summary, and parts of the text that include the most initial occurrences of nouns. Paice (1990) wrote an exhaustive article on the construction of abstracts by means of computer, and reviewed all methods which have been used or proposed for selecting "extract-worthy" sentences from documents.

The concept of "topic sentence" was first used in the field of indexing by Baxendale (1958) who after carrying out a study on a sample of 200 paragraphs found that the first sentence of the paragraph was the "topic sentence" 85 percent of the time, while in another 7 percent it was the last. Extending these results, one can understand that the topic sentence of an article provides a strong indication of its subject matter and that as Lancaster (1991, p.222) stated in accordance with Baxendale's study "the topic sentence was the one judged to be that providing the most information concerning the content". Afterwards Buxton and Meadows (1978) used this concept while they were considering what kind of information is usually taken from each section of an experimental paper to form the abstract. They stated that "the part of the introduction which is most often represented in the abstract is the "topic sentence". This sentence states what exactly is the subject of the study". To clarify their definition of topic sentences, they gave two examples taken from the introduction of articles in two different scientific and non-scientific journals. One of them is the following: "The present study examined children's perception and evaluation of the October Crisis in Quebec", selected from the Journal of Social Psychology, 89 (3), 1973.
According to the computer searches which were made, no-one yet has used or depended on the content of topic sentences as a means to investigate the relevance of title keywords and their representation of the content of their accompanying articles. However, they are currently being used in the production of automatic abstracts where a computer is employed to derive from the text of an item the most representative sentences that express its main ideas. The approaches followed by specialists to the preparation of abstracts via computers have been many and various. It is not appropriate to count them and go through their specifications and characteristics here, as they are already covered by other works such as that of Rush et al. (1971), Craven (1990), Paice (1990), and Lancaster (1991, p.236).

One recent practice of special importance to the work of this chapter is that of Chris Paice who devised and used the term "indicator-phrase" rather than "topic sentence". He uses this concept to identify sentences which bear information about the articles in which they appear. The indicator-phrase method was spoken about and discussed by Paice in 1981 and re-examined by Black and Johnson (1988) who carried out an evaluation study for both the indicator-phrase method due to Paice and the keyword method due to Luhn. Chris Paice (1981) discussed the usefulness of indicator-phrases and their indication of the major topics of their documents. These are structures such as "the principal aim of this paper is to investigate..." and "in the present paper, a method is described for..." which indicate that the remainder of the sentence will tell something of importance within the context of an article.

Paice (1981, p.175) defines such indicator-phrases as "commonly occurring structures which explicitly state that the sentences containing them have something important to say about the subject matter or the message of the document". He regards this method as one out of an armoury of devices
which can be used in the generation of extracts and abstracts by means of computer and his investigations so far " suggest that this may be a particularly powerful and convenient device ".

6.4 Method and samples

6.4.1 The location of topic sentences

The location of topic sentences in the text of an article is not determined by prespecified rules. It is simply on the discretion of authors whether to include them or not, and whether to place them at the beginning of introductions or at the end. This is why the thesis had to consider the whole introduction as well as the objectives part, if it is there, for the selection of a topic sentence, and not to rely on certain paragraphs such as the first or the last paragraph of the introduction of an article.

Topic sentences were taken only from the introduction or objectives sections of articles for more than one reason. First, it is the place where they usually exist. Buxton and Meadows (1978) observed that such statements often occur towards the end of the introduction, though they may be followed by a set of hypotheses to be tested. Hutchins (1978) stated that " we should expect the initial sentences of a text to represent the foundations upon which the writer organizes what he wants to say. The first paragraphs of a text establish, therefore, the ' theme ' for the text as a whole ; they express in essence what the text is going to be about ". Craven (1990) concluded in one of his studies concerning the use of words and phrases in abstracts that " the slightly greater average concentration of abstracts words in the first 200 words of the full text agrees with the advice given to abstractors to pay attention to the first paragraph ". Such advice is mentioned in Maizell et al.(1971, p.27). Second, the importance of the introduction part of an article
was mentioned by Linton (1972) in his simplified style manual for the preparation of articles in social sciences. He stated that "the introduction gives the background of the work and the approaches of the author. It should contain a clear statement of the problem, the scientific questions asked, and the reasons for asking them." It is obvious that the statement of the problem in Linton's quotation is meant to be the topic sentence of the article. The literature is saturated by the style manuals and guides which advise authors who want to prepare papers or articles on the importance of the introduction as they are the first place for them to make contact with their readers. As Hutchins (1978) said "It is here (opening paragraphs) that the reader learns the intentions of the author and how much he is expected to know of the topic already."

As far as objective parts are concerned, they are in fact continuations of the introduction sections. When introductions are thorough, the objective sections constitute a repetition of the topic of the article or disappear. It is only in cases where the introductory parts are solely used for citing the background of the topic and the previous studies that such sections are used to state the aboutness of the article. In this specific study for Arabic articles, authors seem to use the heading "objectives" when the introductions to their articles are not thorough. This was very obvious especially in journals published in the Arab Gulf states.

6.4.2 The matching criteria

Before getting to the level of matching which was used in this process, it is convenient to mention that the criteria of what constituted a keyword were the same as those which were previously used when examining the information content of Arabic titles in chapter four. In this process, only one
matching category was adopted to report the occurrences of title keywords in the topic sentences. There was no need to designate more categories for the sake of knowing whether the matches had been entire, stem, or synonym. As mentioned earlier, the idea behind this chapter was only to see if Arab authors compose unindicative titles for their articles. It was not to investigate the possibility of using keywords of topic sentences as the means to produce printed indexes, nor to be used in Arabic free text retrieval systems. However, this chapter may conclude something important which calls for more investigations in this domain. So, in measuring the title keywords' occurrences in the text of topic sentences, the thesis considered truncation, stem, or synonym matches equivalent to exact or entire matches.

It is worth mentioning here, that phrases such as "this article treats..." and "the aim of this study..." at the beginning of topic sentences are not always directly followed by what the authors will speak about. It was observed that authors sometimes include in their topic sentences what is excluded from their articles. For instance, the phrase "the aim of this article is not to...". In the process of this chapter special consideration was given to this matter and all keywords which appeared after such phrases were not counted as parts of the topic sentences. It is believed that keywords which occur in topic sentences after expressions merely used to indicate a state of exclusion should not appear in titles because titles are normally used to reveal the content of an article.

6.4.3 The recognition of anaphors

In this chapter, attention was paid to ways in which conceptual relations among terms are indicated syntactically and in which concepts can be realised by synonyms. It is well known that concepts and terms in an article
are sometimes referred to by using synonyms and other expressions rather than by repetition of the same terms or concepts. Authors often use other referring expressions, both to improve the cohesiveness of their texts and for stylistic variation. A particularly important aspect of textual cohesion is known as the anaphor. Black (1990) says that anaphor " denotes the use of pronouns in an abbreviated way to objects and events which have been introduced earlier in the discourse ". Liddy (1990) defines anaphor as " the linguistic phenomenon of abbreviated subsequent reference. It is a technique for referring back to an entity which has been introduced with more fully descriptive phrasing earlier in the text. The entity may be an object, a concept, an individual, a process, or state of being. An anaphor refers to this same entity, but with a lexically and semantically attenuated form ". The entity which an anaphor refers to is known as its referent and the preceding expression that first denotes this referent is known as the antecedent of the anaphor.

In this chapter, a topic sentence is said to be tidy if it contains neither anaphors nor other unresolved referential devices which cannot be understood without referring to adjacent sentences. In cases where the sentence was not tidy referent and antecedent sentences were read in order to understand the referential devices found in topic sentences but were not considered in the matching process since they did not exist within the topic sentence boundaries. The resolution of anaphors found in the selected topic sentences was performed manually. On the one hand to do it by computer raises some problems and on the other resolving anaphors by humans seems more accurate. Liddy et al. (1987) and Bonzi and Liddy (1988) mentioned that anaphoric references are easily understood and mentally resolved by human indexers who appear to be able to take abbreviated references into
consideration in constructing appropriate mental representations of text, because in expository texts every new concept is usually introduced to the reader in its fullest, most explicatured form.

An Arabic example and its translation of an anaphor, taken from Arab Affairs (شؤون عربية), 26 (4), 1983 is the following:

تهدف هذه الدراسة أساسًا إلى مقارنة التفكير الاجتماعي الخلدوني بالتفكير الاجتماعي الغربي الحديث. وبالتحديد فإن الهدف هنا هو تبيان أوجه التشابه والإختلاف بين هذين النوعين من الفكر الاجتماعي فيما يخص المحورين التاليين: مفهوم التغيير والتطور الاجتماعي، ثم ما نسميه هنا بالتصنيف السوسيولوجي الذي إستعمله كل من التفكيرين الاجتماعيين لدراسة تحولات المجتمعات.

"This study aims principally to compare the Ibn Khaldun's sociological thinking to the modern Western sociological thinking. Specifically, the aim here is to demonstrate both similarities and discrepancies between these two types with regard to the following two dimensions: The dual concept of sociological change and development and the so-called sociological categorization which was used by both approaches of sociological thinkings to monitor changes in societies ".

In the above text, the anaphoric reference " these two types " is easily understood by human processors to mean "Ibn Khaldun's sociological thinking and the modern Western sociological thinking". The author of the above text can assume that human cognitive processing will replace " these two types " with the words of the phrase it refers to. This process of replacement is known as resolution. Anaphors usually include third person pronouns, nominal demonstratives, indefinites, quantifiers and nominal substitutes.
6.4.4 **Samples and data gathering**

Three constraints restricted the gathering of samples for this study. The first was that only articles with topic sentences located in the prespecified parts were to be taken; the second was the necessity of suppressing all topic sentences which had unresolvable anaphors; and the third, which would rather be called a limitation, was the fact that the British Library and university libraries do not take a wide range of Arabic periodicals for this study to cover more disciplines. To carry out this study, 25 articles each were selected from five different subjects covering the period between 1986 and 1991. All the examined articles were previously put to a reliability test by another researcher to ensure the correctness of their subject classification. The studied subjects comprised agriculture, library and information science, linguistics, philosophy, and sociology. The same indexers in *al-Fihrist* who previously provided the indexing sheets examined earlier in chapter five, were also asked to provide some Arabic articles from the same variety of subjects, but unfortunately they were only able to send samples from five subjects. The fact that not every Arabic article has a topic sentence compelled the study to go through another selection of articles, collected from the available Arabic periodicals in the British Library and university libraries, in order to get the number of samples required. These remaining required articles were selected from other indexes such as the *Index of Dirasat* (دراسات, دراسات) and if taken directly from a periodical, then they were judged for subject according to the content of their topic sentences and not to the content of their titles. Appendix 6.2 shows the names of Arabic journals from which the samples of titles and topic sentences were selected. The aim of this appendix is to show that the sources of samples were not popular
6.5 Results

The results obtained after carrying out the matching process between title keywords and keywords found in topic sentences are spread on five different tables, one for each subject. These tables are shown later in this section as the following: Table 6.1 for agriculture, table 6.2 for library and information science, table 6.3 for linguistics, table 6.4 for philosophy and finally table 6.5 for sociology. Apart from the column which shows the number of the sample article, each single table was divided into five columns. Column A is used for the number of substantive words found in topic sentences, column B is used for the number of substantive words found in titles, column C shows the number of substantive words which matched between topic sentences and titles, and column D shows the relative number of substantive words matched in titles. This column is very crucial for this study to find out the degree of title's representation. Finally the proportion of substantive words which occurred in titles over those which occurred in topic sentences was put in column E. This column was added to facilitate the recognition of differences in the number of keywords found in both topic sentences and titles.
<table>
<thead>
<tr>
<th>No</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.00</td>
<td>4.00</td>
<td>3.00</td>
<td>0.75</td>
<td>0.80</td>
</tr>
<tr>
<td>2</td>
<td>3.00</td>
<td>6.00</td>
<td>3.00</td>
<td>0.50</td>
<td>2.00</td>
</tr>
<tr>
<td>3</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>6.00</td>
<td>3.00</td>
<td>3.00</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>5</td>
<td>18.00</td>
<td>7.00</td>
<td>5.00</td>
<td>0.71</td>
<td>0.38</td>
</tr>
<tr>
<td>6</td>
<td>11.00</td>
<td>10.00</td>
<td>6.00</td>
<td>0.60</td>
<td>0.90</td>
</tr>
<tr>
<td>7</td>
<td>9.00</td>
<td>7.00</td>
<td>6.00</td>
<td>0.85</td>
<td>0.77</td>
</tr>
<tr>
<td>8</td>
<td>8.00</td>
<td>4.00</td>
<td>3.00</td>
<td>0.75</td>
<td>0.50</td>
</tr>
<tr>
<td>9</td>
<td>11.00</td>
<td>5.00</td>
<td>4.00</td>
<td>0.80</td>
<td>0.45</td>
</tr>
<tr>
<td>10</td>
<td>5.00</td>
<td>2.00</td>
<td>2.00</td>
<td>1.00</td>
<td>0.40</td>
</tr>
<tr>
<td>11</td>
<td>13.00</td>
<td>7.00</td>
<td>7.00</td>
<td>1.00</td>
<td>0.53</td>
</tr>
<tr>
<td>12</td>
<td>8.00</td>
<td>5.00</td>
<td>5.00</td>
<td>1.00</td>
<td>0.62</td>
</tr>
<tr>
<td>13</td>
<td>10.00</td>
<td>9.00</td>
<td>7.00</td>
<td>0.77</td>
<td>0.90</td>
</tr>
<tr>
<td>14</td>
<td>5.00</td>
<td>3.00</td>
<td>3.00</td>
<td>1.00</td>
<td>0.60</td>
</tr>
<tr>
<td>15</td>
<td>5.00</td>
<td>6.00</td>
<td>4.00</td>
<td>0.66</td>
<td>1.20</td>
</tr>
<tr>
<td>16</td>
<td>9.00</td>
<td>7.00</td>
<td>7.00</td>
<td>1.00</td>
<td>0.77</td>
</tr>
<tr>
<td>17</td>
<td>8.00</td>
<td>3.00</td>
<td>3.00</td>
<td>1.00</td>
<td>0.37</td>
</tr>
<tr>
<td>18</td>
<td>12.00</td>
<td>5.00</td>
<td>5.00</td>
<td>1.00</td>
<td>0.41</td>
</tr>
<tr>
<td>19</td>
<td>8.00</td>
<td>9.00</td>
<td>3.00</td>
<td>0.33</td>
<td>1.12</td>
</tr>
<tr>
<td>20</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>21</td>
<td>5.00</td>
<td>4.00</td>
<td>4.00</td>
<td>1.00</td>
<td>0.80</td>
</tr>
<tr>
<td>22</td>
<td>6.00</td>
<td>5.00</td>
<td>4.00</td>
<td>0.80</td>
<td>0.83</td>
</tr>
<tr>
<td>23</td>
<td>5.00</td>
<td>3.00</td>
<td>3.00</td>
<td>1.00</td>
<td>0.60</td>
</tr>
<tr>
<td>24</td>
<td>3.00</td>
<td>8.00</td>
<td>3.00</td>
<td>0.37</td>
<td>2.66</td>
</tr>
<tr>
<td>25</td>
<td>3.00</td>
<td>5.00</td>
<td>3.00</td>
<td>0.60</td>
<td>1.66</td>
</tr>
</tbody>
</table>

Table 6.1: number of sample, substantive words in topic sentences (A); substantive words in titles (B); substantive words which matched (C); proportion of substantive words which matched (D); and proportion of substantive words (E) in titles of agriculture.
<table>
<thead>
<tr>
<th>No</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.00</td>
<td>8.00</td>
<td>8.00</td>
<td>1.00</td>
<td>0.66</td>
</tr>
<tr>
<td>2</td>
<td>9.00</td>
<td>4.00</td>
<td>3.00</td>
<td>0.75</td>
<td>0.44</td>
</tr>
<tr>
<td>3</td>
<td>16.00</td>
<td>7.00</td>
<td>5.00</td>
<td>0.71</td>
<td>0.43</td>
</tr>
<tr>
<td>4</td>
<td>10.00</td>
<td>5.00</td>
<td>5.00</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>5</td>
<td>8.00</td>
<td>7.00</td>
<td>6.00</td>
<td>0.85</td>
<td>0.87</td>
</tr>
<tr>
<td>6</td>
<td>4.00</td>
<td>3.00</td>
<td>3.00</td>
<td>1.00</td>
<td>0.75</td>
</tr>
<tr>
<td>7</td>
<td>18.00</td>
<td>7.00</td>
<td>7.00</td>
<td>1.00</td>
<td>0.38</td>
</tr>
<tr>
<td>8</td>
<td>9.00</td>
<td>7.00</td>
<td>5.00</td>
<td>0.71</td>
<td>0.77</td>
</tr>
<tr>
<td>9</td>
<td>5.00</td>
<td>3.00</td>
<td>3.00</td>
<td>1.00</td>
<td>0.60</td>
</tr>
<tr>
<td>10</td>
<td>8.00</td>
<td>5.00</td>
<td>5.00</td>
<td>1.00</td>
<td>0.62</td>
</tr>
<tr>
<td>11</td>
<td>16.00</td>
<td>7.00</td>
<td>7.00</td>
<td>1.00</td>
<td>0.43</td>
</tr>
<tr>
<td>12</td>
<td>6.00</td>
<td>3.00</td>
<td>3.00</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>13</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>14</td>
<td>7.00</td>
<td>5.00</td>
<td>5.00</td>
<td>1.00</td>
<td>0.71</td>
</tr>
<tr>
<td>15</td>
<td>10.00</td>
<td>8.00</td>
<td>7.00</td>
<td>0.87</td>
<td>0.80</td>
</tr>
<tr>
<td>16</td>
<td>8.00</td>
<td>3.00</td>
<td>3.00</td>
<td>1.00</td>
<td>0.37</td>
</tr>
<tr>
<td>17</td>
<td>2.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>18</td>
<td>21.00</td>
<td>6.00</td>
<td>6.00</td>
<td>1.00</td>
<td>0.28</td>
</tr>
<tr>
<td>19</td>
<td>7.00</td>
<td>3.00</td>
<td>2.00</td>
<td>0.66</td>
<td>0.42</td>
</tr>
<tr>
<td>20</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>21</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>22</td>
<td>5.00</td>
<td>7.00</td>
<td>5.00</td>
<td>0.71</td>
<td>1.40</td>
</tr>
<tr>
<td>23</td>
<td>3.00</td>
<td>5.00</td>
<td>2.00</td>
<td>0.40</td>
<td>1.66</td>
</tr>
<tr>
<td>24</td>
<td>8.00</td>
<td>7.00</td>
<td>7.00</td>
<td>1.00</td>
<td>0.87</td>
</tr>
<tr>
<td>25</td>
<td>5.00</td>
<td>8.00</td>
<td>2.00</td>
<td>0.25</td>
<td>1.60</td>
</tr>
</tbody>
</table>

Table 6.2: number of sample, substantive words in topic sentences (A); substantive words in titles (B); substantive words which matched (C); proportion of substantive words which matched (D); and proportion of substantive words (E) in titles of library and information science.
<table>
<thead>
<tr>
<th>No</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.00</td>
<td>4.00</td>
<td>3.00</td>
<td>0.75</td>
<td>0.44</td>
</tr>
<tr>
<td>2</td>
<td>15.00</td>
<td>4.00</td>
<td>4.00</td>
<td>1.00</td>
<td>0.26</td>
</tr>
<tr>
<td>3</td>
<td>5.00</td>
<td>5.00</td>
<td>3.00</td>
<td>0.60</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>7.00</td>
<td>9.00</td>
<td>3.00</td>
<td>0.33</td>
<td>1.28</td>
</tr>
<tr>
<td>5</td>
<td>7.00</td>
<td>5.00</td>
<td>3.00</td>
<td>0.60</td>
<td>0.71</td>
</tr>
<tr>
<td>6</td>
<td>6.00</td>
<td>7.00</td>
<td>4.00</td>
<td>0.57</td>
<td>1.16</td>
</tr>
<tr>
<td>7</td>
<td>8.00</td>
<td>2.00</td>
<td>2.00</td>
<td>1.00</td>
<td>0.25</td>
</tr>
<tr>
<td>8</td>
<td>4.00</td>
<td>3.00</td>
<td>2.00</td>
<td>0.66</td>
<td>0.75</td>
</tr>
<tr>
<td>9</td>
<td>14.00</td>
<td>5.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.35</td>
</tr>
<tr>
<td>10</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>11</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>12</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>13</td>
<td>9.00</td>
<td>8.00</td>
<td>6.00</td>
<td>0.75</td>
<td>0.88</td>
</tr>
<tr>
<td>14</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>15</td>
<td>4.00</td>
<td>3.00</td>
<td>3.00</td>
<td>1.00</td>
<td>0.75</td>
</tr>
<tr>
<td>16</td>
<td>6.00</td>
<td>5.00</td>
<td>5.00</td>
<td>1.00</td>
<td>0.83</td>
</tr>
<tr>
<td>17</td>
<td>7.00</td>
<td>4.00</td>
<td>4.00</td>
<td>1.00</td>
<td>0.57</td>
</tr>
<tr>
<td>18</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>19</td>
<td>5.00</td>
<td>3.00</td>
<td>3.00</td>
<td>1.00</td>
<td>0.60</td>
</tr>
<tr>
<td>20</td>
<td>4.00</td>
<td>2.00</td>
<td>1.00</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>21</td>
<td>3.00</td>
<td>3.00</td>
<td>2.00</td>
<td>0.66</td>
<td>1.00</td>
</tr>
<tr>
<td>22</td>
<td>5.00</td>
<td>4.00</td>
<td>4.00</td>
<td>1.00</td>
<td>0.80</td>
</tr>
<tr>
<td>23</td>
<td>3.00</td>
<td>2.00</td>
<td>2.00</td>
<td>1.00</td>
<td>0.66</td>
</tr>
<tr>
<td>24</td>
<td>5.00</td>
<td>4.00</td>
<td>3.00</td>
<td>0.75</td>
<td>0.80</td>
</tr>
<tr>
<td>25</td>
<td>3.00</td>
<td>3.00</td>
<td>2.00</td>
<td>0.66</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 6.3: number of sample, substantive words in topic sentences (A); substantive words in titles (B); substantive words which matched (C); proportion of substantive words which matched (D); and proportion of substantive words (E) in titles of linguistics.
<table>
<thead>
<tr>
<th>No</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.00</td>
<td>6.00</td>
<td>1.00</td>
<td>0.16</td>
<td>2.00</td>
</tr>
<tr>
<td>2</td>
<td>5.00</td>
<td>2.00</td>
<td>2.00</td>
<td>1.00</td>
<td>0.40</td>
</tr>
<tr>
<td>3</td>
<td>6.00</td>
<td>3.00</td>
<td>1.00</td>
<td>0.33</td>
<td>0.50</td>
</tr>
<tr>
<td>4</td>
<td>3.00</td>
<td>2.00</td>
<td>2.00</td>
<td>1.00</td>
<td>0.66</td>
</tr>
<tr>
<td>5</td>
<td>12.00</td>
<td>3.00</td>
<td>2.00</td>
<td>0.66</td>
<td>0.25</td>
</tr>
<tr>
<td>6</td>
<td>11.00</td>
<td>9.00</td>
<td>7.00</td>
<td>0.77</td>
<td>0.81</td>
</tr>
<tr>
<td>7</td>
<td>3.00</td>
<td>2.00</td>
<td>2.00</td>
<td>1.00</td>
<td>0.66</td>
</tr>
<tr>
<td>8</td>
<td>6.00</td>
<td>3.00</td>
<td>3.00</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>9</td>
<td>6.00</td>
<td>2.00</td>
<td>2.00</td>
<td>1.00</td>
<td>0.33</td>
</tr>
<tr>
<td>10</td>
<td>4.00</td>
<td>2.00</td>
<td>2.00</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>11</td>
<td>7.00</td>
<td>5.00</td>
<td>5.00</td>
<td>1.00</td>
<td>0.71</td>
</tr>
<tr>
<td>12</td>
<td>4.00</td>
<td>2.00</td>
<td>2.00</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>13</td>
<td>2.00</td>
<td>4.00</td>
<td>2.00</td>
<td>0.50</td>
<td>2.00</td>
</tr>
<tr>
<td>14</td>
<td>16.00</td>
<td>7.00</td>
<td>1.00</td>
<td>0.14</td>
<td>0.43</td>
</tr>
<tr>
<td>15</td>
<td>4.00</td>
<td>6.00</td>
<td>1.00</td>
<td>0.16</td>
<td>1.50</td>
</tr>
<tr>
<td>16</td>
<td>18.00</td>
<td>12.00</td>
<td>7.00</td>
<td>0.58</td>
<td>0.66</td>
</tr>
<tr>
<td>17</td>
<td>7.00</td>
<td>6.00</td>
<td>6.00</td>
<td>1.00</td>
<td>0.85</td>
</tr>
<tr>
<td>18</td>
<td>6.00</td>
<td>6.00</td>
<td>6.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>19</td>
<td>15.00</td>
<td>3.00</td>
<td>3.00</td>
<td>1.00</td>
<td>0.20</td>
</tr>
<tr>
<td>20</td>
<td>11.00</td>
<td>8.00</td>
<td>2.00</td>
<td>0.25</td>
<td>0.72</td>
</tr>
<tr>
<td>21</td>
<td>15.00</td>
<td>10.00</td>
<td>10.00</td>
<td>1.00</td>
<td>0.66</td>
</tr>
<tr>
<td>22</td>
<td>5.00</td>
<td>4.00</td>
<td>3.00</td>
<td>0.75</td>
<td>0.80</td>
</tr>
<tr>
<td>23</td>
<td>5.00</td>
<td>5.00</td>
<td>3.00</td>
<td>0.60</td>
<td>1.00</td>
</tr>
<tr>
<td>24</td>
<td>5.00</td>
<td>4.00</td>
<td>2.00</td>
<td>0.50</td>
<td>0.80</td>
</tr>
<tr>
<td>25</td>
<td>6.00</td>
<td>5.00</td>
<td>3.00</td>
<td>0.60</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Table 6.4: number of sample, substantive words in topic sentences (A); substantive words in titles (B); substantive words which matched (C); proportion of substantive words which matched (D); and proportion of substantive words (E) in titles of philosophy.
<table>
<thead>
<tr>
<th>No</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.00</td>
<td>6.00</td>
<td>1.00</td>
<td>0.16</td>
<td>3.00</td>
</tr>
<tr>
<td>2</td>
<td>6.00</td>
<td>4.00</td>
<td>2.00</td>
<td>0.50</td>
<td>0.66</td>
</tr>
<tr>
<td>3</td>
<td>10.00</td>
<td>4.00</td>
<td>4.00</td>
<td>1.00</td>
<td>0.40</td>
</tr>
<tr>
<td>4</td>
<td>12.00</td>
<td>7.00</td>
<td>6.00</td>
<td>0.85</td>
<td>0.58</td>
</tr>
<tr>
<td>5</td>
<td>15.00</td>
<td>5.00</td>
<td>5.00</td>
<td>1.00</td>
<td>0.33</td>
</tr>
<tr>
<td>6</td>
<td>8.00</td>
<td>4.00</td>
<td>4.00</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>7</td>
<td>5.00</td>
<td>5.00</td>
<td>3.00</td>
<td>0.60</td>
<td>1.00</td>
</tr>
<tr>
<td>8</td>
<td>5.00</td>
<td>2.00</td>
<td>2.00</td>
<td>1.00</td>
<td>0.40</td>
</tr>
<tr>
<td>9</td>
<td>7.00</td>
<td>7.00</td>
<td>4.00</td>
<td>0.57</td>
<td>1.00</td>
</tr>
<tr>
<td>10</td>
<td>11.00</td>
<td>6.00</td>
<td>4.00</td>
<td>0.66</td>
<td>0.54</td>
</tr>
<tr>
<td>11</td>
<td>12.00</td>
<td>4.00</td>
<td>4.00</td>
<td>1.00</td>
<td>0.33</td>
</tr>
<tr>
<td>12</td>
<td>12.00</td>
<td>7.00</td>
<td>5.00</td>
<td>0.71</td>
<td>0.58</td>
</tr>
<tr>
<td>13</td>
<td>17.00</td>
<td>6.00</td>
<td>4.00</td>
<td>0.66</td>
<td>0.35</td>
</tr>
<tr>
<td>14</td>
<td>6.00</td>
<td>9.00</td>
<td>6.00</td>
<td>0.66</td>
<td>1.50</td>
</tr>
<tr>
<td>15</td>
<td>20.00</td>
<td>5.00</td>
<td>5.00</td>
<td>1.00</td>
<td>0.25</td>
</tr>
<tr>
<td>16</td>
<td>9.00</td>
<td>6.00</td>
<td>6.00</td>
<td>1.00</td>
<td>0.66</td>
</tr>
<tr>
<td>17</td>
<td>8.00</td>
<td>4.00</td>
<td>3.00</td>
<td>0.75</td>
<td>0.50</td>
</tr>
<tr>
<td>18</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>19</td>
<td>12.00</td>
<td>8.00</td>
<td>6.00</td>
<td>0.75</td>
<td>0.66</td>
</tr>
<tr>
<td>20</td>
<td>9.00</td>
<td>7.00</td>
<td>6.00</td>
<td>0.85</td>
<td>0.77</td>
</tr>
<tr>
<td>21</td>
<td>4.00</td>
<td>2.00</td>
<td>2.00</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>22</td>
<td>4.00</td>
<td>5.00</td>
<td>3.00</td>
<td>0.60</td>
<td>1.25</td>
</tr>
<tr>
<td>23</td>
<td>3.00</td>
<td>7.00</td>
<td>3.00</td>
<td>0.42</td>
<td>2.33</td>
</tr>
<tr>
<td>24</td>
<td>3.00</td>
<td>3.00</td>
<td>2.00</td>
<td>0.66</td>
<td>1.00</td>
</tr>
<tr>
<td>25</td>
<td>3.00</td>
<td>2.00</td>
<td>2.00</td>
<td>1.00</td>
<td>0.66</td>
</tr>
</tbody>
</table>

Table 6.5: number of sample, substantive words in topic sentences (A); substantive words in titles (B); substantive words which matched (C); proportion of substantive words which matched (D); and proportion of substantive words (E) in titles of sociology.
Since the number of substantive words in a topic sentence varied from one article to another, a measure of title content should not be biased by the length of the topic sentences. Standardized comparisons among this variety of subjects was possible by expressing matches of substantive words as the ratio of the number of actual matches between keywords in the topic sentence and keywords in the title over the number of substantive words in the title. This ratio is expected to have in each case a maximum score of 1, and a minimum of zero. For example, an Arabic article with four title keywords, three of which matched words in its topic sentence, had a relative match for the topic sentence of 3/4 or 0.75.

To discuss the results shown in the previous tables, two more tables had to be added to this section. These were tables 6.6 which organizes the examined subjects according to their mean (M) number of the relative matching scored for each of them. The other one was table 6.7 which displays the same examined subjects according to the value of their mean (M) number of the absolute matching in titles (C) and it as well displays, the mean and standard deviation number of substantive words found in Arabic titles (B) and topic sentences (A).
### Table 6.6: Mean and Standard Deviation Value of the Relative Matching in Titles of Each of the Examined Subjects, Arranged According to Their Mean Values.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Library&amp;Infor.Sc.</td>
<td>0.88</td>
<td>0.20</td>
</tr>
<tr>
<td>Agriculture</td>
<td>0.82</td>
<td>0.21</td>
</tr>
<tr>
<td>Linguistics</td>
<td>0.79</td>
<td>0.26</td>
</tr>
<tr>
<td>Sociology</td>
<td>0.78</td>
<td>0.23</td>
</tr>
<tr>
<td>Philosophy</td>
<td>0.72</td>
<td>0.32</td>
</tr>
</tbody>
</table>

### Table 6.7: Mean (M) and Standard Deviation (S.D.) Number of the Absolute Matching in Titles (C); Substantive Words in Titles (B); and Substantive Words in Topic Sentences (A), Arranged According to the Mean Values of Their Absolute Matching.

<table>
<thead>
<tr>
<th>Subject</th>
<th>(C) M</th>
<th>(C) S.D.</th>
<th>(B) M</th>
<th>(B) S.D.</th>
<th>(A) M</th>
<th>(A) S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Library&amp;Infor.Sc.</td>
<td>4.52</td>
<td>1.92</td>
<td>5.28</td>
<td>1.99</td>
<td>8.40</td>
<td>4.88</td>
</tr>
<tr>
<td>Agriculture</td>
<td>4.24</td>
<td>1.48</td>
<td>5.48</td>
<td>2.10</td>
<td>7.44</td>
<td>3.60</td>
</tr>
<tr>
<td>Sociology</td>
<td>3.84</td>
<td>1.52</td>
<td>5.16</td>
<td>1.89</td>
<td>8.28</td>
<td>4.73</td>
</tr>
<tr>
<td>Linguistics</td>
<td>3.32</td>
<td>1.65</td>
<td>4.36</td>
<td>1.91</td>
<td>6.12</td>
<td>3.14</td>
</tr>
<tr>
<td>Philosophy</td>
<td>3.20</td>
<td>2.31</td>
<td>4.84</td>
<td>2.73</td>
<td>7.40</td>
<td>4.58</td>
</tr>
</tbody>
</table>
According to Table 6.6 the titles constructed by Arab authors in five different fields provide an article representation which includes a mean ranging from 72% to 88% of the content reflecting keywords in topic sentences. The highest mean of matches was found in library and information science and the lowest was found in philosophy. Between these two fields, the subject of agriculture scored a mean match of 82% followed by linguistics 79% and sociology 78%.

6.6 Discussion

The purpose of including topic sentences in an author's works is to give readers a clue to the aboutness of such works, so they often can estimate the degree of importance of these works to their researches and interests. Another reason, although secondary but which may be effective, is that authors are in fact drawing the boundaries for their works so they can keep themselves in context. Whether this latter minor reason is true or false needs separate investigations.

The language in which an article is written has no effect on the purpose and the decision of whether to include a topic sentence or not. This is why most of the phrases found at the beginning of Arabic topic sentences and English topic sentences are almost the same in translation. Appendix 6.1 which includes all phrases of Arabic topic sentences found in this study translated into English, reflects this notion. An English researcher, who is familiar with the type of topic sentences used in English literature, can easily realise that the translated Arabic phrases are not different from those which usually appear in English topic sentences. Yet the language affects other aspects such as the structure of topic sentences which affects the sorting and the retrieval processes, but such problems are beyond the scope of this
chapter since it is not yet suggested whether using keywords found in topic sentences to generate Arabic printed indexes would be more relevant. This idea may be a good proposal for a new study in Arabic, especially after it was found in this study that the number of articles which had more keywords in their titles than in their topic sentences was relatively small. Out of 125 articles, only 17 had more keywords in the titles (13.6%). This can be easily observed by looking at column E. This reveals that agriculture had 5 articles, library and information science had 3, linguistics had 2, philosophy had 3 and finally sociology had 4 articles.

Since it was observed during the data gathering for this chapter that only a few articles had abstracts, this study suggests that more attention should be paid to the existence of Arabic topic sentences in the introduction or objectives parts of Arabic articles. The existence of topic sentences will no doubt play an important role in revealing the content of their articles. This will, in turn, enhance the proposal of examining the reliance on keywords in topic sentences to produce automatic derivative Arabic printed indexes. The production of such indexes and the study of its feasibility necessitates that every single article published in academic Arabic journals has a topic sentence. This study found that an average of two in every five Arabic articles had no apparent topic sentences in their introduction or objectives parts. Since the problems of keywords relevance in topic sentences for retrieval and other aspects involved in the production of printed indexes were beyond the interests of this chapter, in matching Arabic title keywords with topic sentences, all truncation and synonym matches were considered equivalent to strict matches. Sparck Jones and Kay (1973) observed previously that truncation operations, such as suffix stripping, provide an adequate representation of the original keywords.
All the results which were shown in the previous section stem from the assumption that the topic sentence of an Arabic article reflects the content of the article itself and that against it one can measure the content-bearing quality of the article's title. Before answering the question of whether Arabic titles in the above examined fields are representative or not, one more question had to be considered.

*How many matches are needed between substantive words in titles and substantive words in topic sentences of their accompanying articles, for these titles to be considered representative?*

It would not be easy to answer such a question quantitatively by requesting a certain number of matches to occur between title keywords and keywords in topic sentences. However, if the topic sentence uses a certain number of substantive words to represent the subject matter of its accompanying article, then the title of this article ought to contain a certain percentage of these substantive words.

In previous studies concerned with titles' informativeness, Tocatlian (1970) "assumed that titles with three or less substantive words are most likely uninformative". Afterwards, Peritz (1984) counted a title as informative "if it conveyed at least some general idea of the paper's content, without recourse to other sources of information such as the abstract, the journal title, or the paper itself. In other words, the title was considered informative if, upon reading the title, one felt that one knew what it was about ".

The criteria of judging titles in the two aforementioned studies are not used in this study which is concerned with titles' representativeness. On the one hand, Tocatlian's criterion seemed arbitrary since he depended on an arbitrary definition of an informative title. He mentioned that if "six
substantive words " had been stated arbitrarily in the literature as an average number to make an informative title, then " four to eight substantive words could be taken as the range ". On the other hand, Peritz's categorization of informative and non-informative titles was not quantitative. For instance, she did not suggest whether one, two or three informative words in a title do or do not give " some general idea of the paper's content ". After all, it might be possible to depend on personal feelings or on the number of keywords to judge the informativeness of titles but how can their representativeness be measured or tested without comparing them with the text which they are supposed to represent ? It is quite obvious that an informative title is not necessarily representative and that the degree of representativeness can change between one title and another according to the number of matches between their keywords and the keywords of their comparable texts, in this study the topic sentences.

In this study only titles which had no matches between their substantive words and those found in their accompanying topic sentences were categorized as non-representative. Other titles were described as representative although the degree of representativeness differs from one title to another. Out of 125 titles studied in five different subjects, only one article title had no matches. This title was found in linguistics. Table 6.8 shows the sample studied in this chapter divided into two categories; representative and non-representative along with their percentages.

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representative titles</td>
<td>124</td>
<td>99.2%</td>
</tr>
<tr>
<td>Non-representative titles</td>
<td>1</td>
<td>0.8%</td>
</tr>
<tr>
<td>Total</td>
<td>125</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 6.8 : number and percentage of representative and unrepresentative titles in the five selected subjects, in accordance with Peritz (1984) criterion.
Table 6.6 revealed the variability of title representativeness among Arabic disciplines as well as the relatively high percentage of representation that each of the examined subjects had, when it was investigated according to the criteria of this study. Although philosophy was the only subject which had less than 75% relative matches between keywords in topic sentences and keywords in titles, a glance at all other percentages shows that the library and information science articles tend to have more representative content in their titles than do agriculture articles and significantly greater representative content than do other articles in linguistics, sociology and philosophy according to the occurrence of Arabic title keywords in their accompanying topic sentences. Linguistics and sociology resembled each other and there was no significant difference between the content in their titles and the content in philosophy article titles. Column C in table 6.7 showed that library and information science also had the highest absolute number of matches while philosophy had the lowest.

For the field of library and information science to score the highest percentage of content representation in this group of subjects is not totally strange. Generally specialists in this field are much more aware of the advantages of assigning representative titles to their articles. Suffice it to mention that they are familiar with how indexers interpret articles and since they constantly use bibliographic tools in their professions, apparently they know what parts of an article are most significant for indexers. What was quite surprising for this study was to find article titles of non-scientific disciplines in Arabic literature scoring a considerable percentage of content representation.

Assuming that the topic sentence of an article is the part where authors
of articles express the subject matter of their articles, this study revealed that informativeness is something different from representativeness. To conclude that titles of certain subjects are representative or relevant by merely counting the keywords they contain is inadequate. Perhaps it is appropriate to mention the linguistics' sample title which did not have any matches and which according to the conventions of this study was not representative. This Arabic title, taken from the Arabic Journal of Humanities (الجلة العربية للعلوم الإنسانية), 11 (22), 1986 and its English translation are the following:
المجز اللغوي كسلوك لغوي للإنسان المغربي المغلوب

"Language ambiguity as a distinguished behaviour of the Moroccan underdog citizen ".

The topic sentence of this article and its English translation are the following:

"This study aims at demonstrating that the Franco-Arab phenomenon can only be fully explained from the imperialist perspective as being one symptom of the influence of French colonization on these societies during the nineteenth and the twentieth centuries ".

When it was referred to the text of the article, one of the most important concepts was found to be " the Franco-Arab phenomenon ". This occurred in seven out of eleven section headings. Also, it was found that the article was not specifically about Morocco, as is understood from the title. It referred rather to all former French colonies in North Africa. The topic sentence has both concepts named as " Franco-Arab phenomenon " and " French
colonization ". In this case it was felt that the topic sentence was a good representation of the content of the article, whilst the title was not.

According to the method used in this study, longer titles or titles which contain more substantive words are not necessarily more representative: In table 6.7, column B and column C, which respectively display the mean and standard deviation numbers of substantive words found in titles and the mean and standard deviation numbers of the absolute matching in titles of the five selected disciplines, reveal that titles in agriculture which had the largest number of keywords (5.48 substantive words) were not the most representative. They also revealed that titles in linguistics, although the shortest (4.36 substantive words), were not the least representative. The order according to their absolute number of matches in titles is: library and information science, agriculture, sociology, linguistics and philosophy.

6.7 Conclusion

The samples used in this study belong to particular subjects and a particular style of writing. All were written in Arabic and taken from academic Arabic journals. This study found that a mean ranging from 70% to 88% of title keywords occurred in the topic sentences. Assuming that keywords in topic sentences reflect the content of their accompanying articles, these results are more than encouraging towards the dependence on title substantive words to produce Arabic title keyword printed indexes. The high percentage of similarity found between the Arabic title keywords and the keywords in topic sentences indicates that Arab indexers, when considering the title keywords in the process of indexing Arabic articles, are in fact depending on representative titles. However, indexers seeking clues to the content of articles in the five examined subjects would find that titles of
articles in library and information science and agriculture are more useful than those in linguistics, sociology and philosophy. As far as Arabic topic sentences are concerned, this study revealed that only 60% of the Arab authors include topic sentences in the introduction and objectives parts of their articles.

Although the results of this study do not contradict the encouraging results which were obtained previously and prove that Arabic titles are representative, this does not mean that the production of Arabic title-derivative printed indexes is a very straightforward process. There are still other issues to be investigated. For instance the reaction of users and the problems of retrieving from Arabic printed indexes have not been investigated yet. These two issues can be very important towards deciding whether producing such indexes is useful in Arabic. The following two chapters are concerned with examining these two issues before reaching the final conclusion of this thesis.
chapter seven

Producing an Arabic title-derivative printed index and users' evaluation

Having seen that Arabic article titles are representative and reflect to a large extent the contents of their corresponding articles, it is now time to approach other practical aspects of the subject of relying on keywords of Arabic titles to produce printed indexes. Therefore, the aim of this chapter is to produce a pilot title-derivative printed index for some Arabic articles. Then it will study the users' evaluation to title keyword indexing through a comparison between that index and its equivalent conventional printed index produced for the same set of articles. Moreover, since title-derivative printed indexes for English script materials are generally produced automatically, this chapter also investigates the issue of automatically generating Arabic printed indexes and the problems involved.
7.1 Bibliographic information systems and Arabization

Over the last three decades many academic and commercial Arab institutions inside and outside the Arab countries have introduced Latin alphabet bibliographic systems to their libraries and information centres. Al-'Arabi (1982) mentioned that the National Agricultural Documentation Centre in Tunisia has been relying on computers since 1975. However, because their collections include large quantities of Arabic material, librarians and information specialists have raised the issue of making the available systems compatible with the Arabic language, so they can be used to accommodate and handle bibliographic operations for records of Arabic materials. Since then, the search for relevant bibliographic systems which can take into account the peculiar needs and features of the Arabic language have started on different levels with obvious lack of cooperation and coordination. Khurshid (1992) stated that the plan for automating bibliographic operations and processes at the King Fahd University of Petroleum and Minerals Library "began in 1975, when five systems were investigated by the library administration".

The process of moving to computerized Arabic bibliographic systems has not been easy. Aman (1984), Booth et al. (1986), and Ashoor (1989a) wrote extensively about the process and its requirements. They also mentioned the problems which faced and are still facing the Arabization programme including the absence of a standard coding system for Arabic characters and many other technical and linguistic problems imposed by the peculiarity of the Arabic language. As far as the coding system is concerned, it was only in 1982 that a breakthrough was made when ASMO (Arab Organization for Standardization and Metrology) 449 which is compatible
with the seven bits standard became the standard coding system for the Arabic language, and afterwards the release of ASMO 708 which is compatible with the eight bits standard. Al-'Akhras (1982), al-Najdawi (1982), Clews (1988a) and Ashoor (1989b) described the problems concerned and the attempts made in the past by some Arab institutions to devise a unified coding system for Arabic characters and the efforts behind the creation of ASMO 449. Jaffal (1993) mentioned the latest Arabic coding system ASMO 708 and confirmed that it has overcome the few problems of the previous system with regards to Arabic / Latin information exchange.

According to the available literature on Arabization, the earliest attempt of Arabizing a bibliographic information storage and retrieval system was initiated by ALDOC (Arab League DOcumentation Centre) in 1980, as a part of a cooperative agreement with IDRC (International Development Research Centre) to Arabize MINISIS (al-Qasimi,1987). ALDOC assumed full responsibility for the Arabization of MINISIS, but the process faced various hardware and software problems including the absence of a standard Arabic characters coding system and a bilingual terminal that has a complete Arabic set. Therefore, the Arabization of the system was delayed and took almost three years before it was implemented.

As far as other early Arabization activities are concerned, Ashoor (1989a) cited two attempts separately carried out by KSU (King Sa‘ud University) with DOBIS/LIBIS and by some Kuwaiti governmental bodies with STAIRS. Both systems run on IBM computer machines, whereas MINISIS runs on any of the HP3000 series of minicomputers. Ashoor and Chaudhry (1992) evaluated MINISIS and DOBIS/LIBIS as they are considered to be the most fruitful of the early Arabization attempts and provided a list of bilingual bibliographic software which is currently in use in
Ashoor and Chaudhry (1992) scanned a large number of directories of libraries and information institutions, conference proceedings and other relevant recent journal articles to develop a list of 15 institutions which seemed "to have reasonable involvement in performing bibliographic activities on bilingual materials". Out of the 15 institutions, only 10 turned out to be using computers and of the 10 only two institutions are using inhouse developed programs.

The extension of Ashoor and Chaudhry's study to cover more Arab institutions in the Arab countries is not necessary to confirm the fact that few institutions chose to start their bilingual programs from scratch and the fact that the institutions which chose to Arabize existing Latin systems did not cooperate with each other but preferred to work independently. Moreover it is realised that most of these institutions are governmental bodies in the rich Gulf states or belong to regional or international organizations such as ALDOC.

However, the situation in the current decade is different and the use of computer programs which handle Arabic script is not as difficult as in the early application of bibliographical operations on Arabic materials. Currently most computer manufacturers and systems houses consider the multilingual aspects of their products. IBM, Apple, Microsoft and al-'Alamiyyah have systems offering Arabic and other alphabets character support. In fact, the provision of multilingual script compatible operating systems by such computer manufacturers has made it possible to modify packages whose native mode is not Arabic into an Arabic working environment. Kilany and Sharif (1988) provided a list of some UK-based Arabization system houses. Barkatullah (1988) cited some word processing packages and database
management systems which are Arabic script compatible. Tresman (1991) mentioned some computer products that accept the Arabic language in his source guide for multilingual and foreign language products for IBM PCs and compatibles.

7.2 The use of computers to produce Arabic printed indexes

According to the literature which this thesis surveyed, only two institutions use the computer facilities to produce their current Pan-Arab multiperiodical printed indexes. Dilmun for Publishing uses the Dilmun system to produce two indexer-assigned keyword printed indexes: the Islamic Index and the Palestine Record. Ubaidy (1992) said that the Dilmun system is an original program created for Dilmun and covers most of the functions related to indexing and index manipulation. It runs on IBM computers and IBM system compatibles. The other institution is the Information and Documentation Centre at the Gulf Cooperation Council. It produces the Analytical Index of the Arabic and Foreign Periodicals and Bulletins at the Information and Documentation Centre of the Gulf Cooperation Council Secretariat General using the Arabized version of STAIRS. Ashoor (1989a) stated that "the Arabized version of STAIRS removes some restrictions in the naming conventions, and character set usage and extends the system in order to make it usable by personnel having little knowledge of the English language and the program has been designed to accommodate both English and Arabic records in one database...".

The use of a computer to produce Arabic printed indexes for periodical
articles is not restricted to the production of current printed indexes. Computer facilities have been also used to produce printed indexes for individual journals as well as for retrospective periodical articles. For instance, the King Fahd National Library used MINISIS to produce the Index of Saudi Periodicals (كشاف الدوريات السعودية) as well as the Analytical Index of King 'Abd al-'Aziz University Journals (الكشاف التحليلي لمجلات جامعة الملك عبد العزيز) and the Index of Articles in Saudi Periodicals Specialized in Libraries and Information (الكشاف بالمقالات في الدوريات السعودية المتخصصة في المكتبات والمعلومات). The Yearly Index of al-Faysal Journal (الكشاف السنوي لمجلة الفيصل) is prepared by means of MINISIS. ALECSO (the Arab League Educational, Cultural, and Scientific Organization) produced the Bibliographical Guide to the Arabic Writings on Information (الدليل الببليوغرافي للإنتاج الفكري العربي في) (1985-1981) in 1987 using a computer program, apparently MINISIS. 'Ibn Khamis (1987, p.5) stated in the introduction of that guide that it was prepared by means of computer but did not mention the software used. King 'Abd al-'Aziz Al Sa'ud Foundation for Islamic Studies and Humanities (مؤسسة الملك عبد العزيز آل سعود للدراسات الإسلامية والإنسانية) in Casablanca depends on MINISIS facilities to produce its quarterly new acquisitions bulletin the Bibliography of the Islamic Maghrebin (الببليوغرافيا الغرب الإسلامي).

In fact the absence of sufficient introductory guides to both Arabic bibliographic software and Arabic printed indexes makes the job of counting printed indexes which use computers tiresome and tedious. There may be some other printed indexes produced by computer which are not available in the market or even not declared officially. Moreover, there may be some software packages which can be used to produce Arabic printed indexes, but
no one is using them for that purpose or has ever heard about them. Al-Hayat Information Center is using OMNIS 3.3 to index the *al-Hayat* (الحياة) daily newspaper. Xperts Librarian which is used by al-Furqan Library as a multilingual library management software can easily be transformed to produce printed indexes for Arabic journal articles. In his study concerning libraries and information in the Middle East, Francis (1993, p.112) mentioned other systems which are in use in single libraries. These systems include CLSI, VTLS, DATA-TREK, and SOUTRON. Apparently these systems accept Arabic script, since they were mentioned along with other Arabized systems in use in the Arab countries such as DOBIS/LIBIS and MINISIS.

It is observed that the number of current Arabic printed indexes is relatively small. Therefore, the names of these indexes will be repeatedly mentioned below, while the thesis is trying to treat the aspects of producing title word printed indexes for Arabic articles by means of computer. It is also observed that the indexing procedure followed in preparing these indexes is human-assigned indexing, involving the representation of subject matter by means of subject headings and keywords selected from some form of controlled vocabulary. However as will be seen, some of these indexes use computer techniques in their formatting.

### 7.3 Title keyword derivation techniques and the Arabic language

The automatic production of title keyword indexes requires the availability of suitable software to carry out the keyword generation and then the formatting process according to a prespecified technique. One of the best known techniques for generating title keyword indexes by means of computer is the rotation technique. The most popular type of rotated printed index is KWIC (the KeyWord-In-Context) (see section 3.4.1.A). Another well known
type which uses the same principle as KWIC but with modifications in its final format, is KWOC (the KeyWord-Out-of Context technique) (see section 3.4.1.B). The other familiar technique used in the production of title-derivative printed indexes is referred to as the title keyword permutation technique or title keywords coordination, like the PSI and Pandex indexes (see section 3.4.3.A and 3.4.3.C).

7.3.1 The rotation technique and the Arabic language

None of the previously mentioned Arabic printed indexes depends on a computer to generate automatically its main entries or keywords. The rotated keywords in both of Dilmun's publications the Islamic Index (الكشف على الإسلام) and Palestine Records (الذاكرة الفلسطينية) are manually selected by indexers from the full texts of the corresponding items and sometimes from the thesaurus of al-Jami'a (الجامعة), and are then rotated by means of computer and put into their indexing positions (see respectively sections 2.3.2.B.4 and 2.3.2.B.5 and figures 2.6 and 2.7).

The keyword rotation technique used by Dilmun's bibliographic publications was not completely new to Arab information scientists who had seen such a technique applied before, in other Arabic bibliographic tools. Arab information scientists first found out about the keyword rotation technique in Winter 1986 when it was adopted as the technique to be used in the production of the Index to Events section inside the quarterly trilingual bibliographic journal Panorama of Events (حالييات حاليات), which used to be published by Publishing and Marketing House in Lebanon. Panorama of Events was a bibliography of news and documents published in various Arabic periodical sources on Lebanese affairs and the Middle East conflict. It started in 1977 and ceased publication in 1987. The indexing keywords for
each news story were humanly assigned and sometimes manually taken from
the news abstracts' section and then rotated throughout the index by means of
a specially made computer program. A sample of the Panorama of Events'
Index to Events is shown in figure 7.1.

7.3.2 The permutation technique and the Arabic language

As for the permutation and the keywords coordination techniques, this
research did not come across any Arabic bibliographic tools which rely on
such methods in their automatic or manual production. It is believed that
these techniques have never been tried before on Arabic literature, although
there are some Arab publishers and information scientists who are confused
in describing rotated indexes as permuted. In fact, Dilmun's indexes are being
produced using the rotation technique with a final format more or less similar
to indexes which use the permutation technique like the PSI. For instance, in
the Islamic Index (الكشف الإسلامي) the number of access points is equal to
the number of the assigned keywords, whereas in the PSI, which is a
permuted index, each keyword is permuted by means of a computer package
to generate all possible pairs of keywords between primary and co-terms. As
was quoted from Garfield (1976) in chapter three, section 3.4.3.A, after
suppressing the insignificant words, this usually generates about 42
words-pairs for the typical seven word title ".

7.4 The title-derivative index types and the Arabic language

Under this section the indexing types of KWIC, KWOC, and the PSI
will be approached, in order to see whether such types are known to Arab
users. It will also be seen whether such types have been employed in
producing tools other than printed indexes in the Arab countries.
7.4.1 The KWIC type and the Arabic language

The first KWIC index to be shown in the Arabic language consisted of a single page sample index to 17 Arabic articles on librarianship. It was prepared manually by Kasem (1985), for an article about the possibility of relying on titles to produce Arabic KWIC indexes (كشف الكلمات المفتاحية داخل السياق). At that time Kasem (1985) mentioned that the production of such indexing types is possible with computer facilities. However, he depended on a typewriter to prepare and print the sample index. This sample index is shown in figure 7.2.

The other Arabic KWIC type product which was identified was a rotation for an Arabic medical terminology list prepared jointly by the Arab Centre for Medical Literature and the National Computer and Microfilm Centre in Kuwait. Salem and el-Haddad (1990) stated that "the permutation list was of great help at the stage of technical revision of the terms". The aim was to set up a bilingual medical terminology database. A sample of that list is shown in figure 7.3. According to Salem and el-Haddad (1990), the production of that list was completely automatic, with human intervention only during the data input to the system. This process indicated that the rotation technique which leads to the KWIC index type is possible for Arabic characters.
Figure 7.1: sample of Index to Events Section in the Panorama of Events Journal
Figure 7.2: Kasem's manual KWIC index
Figure 7.3: Salem's KWIC list of Arabic medical terminology
7.4.2 The KWOC type and the Arabic language

The KWOC type has never been tried on titles of Arabic articles, but it was used to produce a KWOC type index (كشاف الكلمات المفتاحية خارج السياق) to the Arabic descriptors included in the thesaurus of al-Jami‘a (الجامعة) which was published by ALECSO in 1987. Apparently, the Arabized version of MINISIS was used to carry out this process. It is observed that the program succeeded in automatically stripping the definite article ‘al (ال) from every descriptor containing it, and also in suppressing some stopwords such as the separable preposition min (من) from the descriptor الاستشعار من بعد التضARDS (remote sensing) and the separable preposition ‘ila (إلى) from the descriptor الإعادة إلى بلد المنشأ (restitution to country of origin). On the other hand, the system did not succeed in stripping other inseparable prepositions (prefixes), such as the preposition letter l (ل) from the word التصنيف الدولي للتصاميم الصناعية (international classification of industrial designs) and the preposition letter b (ب) from the word بالجفاف (bil-jafaf, for drought) in the descriptor المناطق المعرضة للجفاف (drought-stricken areas).

The result of this unsuccessful stripping process was that the word للتصاميم (lil-tasmim, of designs) appeared under the letter l (ل) within the alphabetical sorting of descriptors and not under the letter t (ت). Also, the word بالجفاف (bil-jafaf, for drought) came under the letter b (ب) and not under the letter j (ج). As for the success in stripping the inseparable preposition letter w (و), it is believed that this was due to the fact that every such letter was keyboarded as a separate character from its following word during the input stage and then considered as a stopword. Otherwise, the
system would have suppressed the other two inseperable prepositions ب (ب) and ل (ل) which repeatedly appeared in the core of the KWOC index of *al-Jami'a* thesaurus joined to their following descriptors as in the word بالحاسب (bil-hasub, with computer).

It seems that the automatic stripping of the Arabic inseparable prepositions from their joined words is not yet an easy process. These prepositions most of the time constitute real parts of their joined words as in the descriptors اللغات (al-lughat, languages), where the letter ل (ل) is a part of the word and not a preposition. The same is the case for the descriptor البطالة (al-batalah, unemployment), where the letter ب (ب) is a real part of the descriptor. Therefore, it is believed that human intervention was necessary for this KWOC type to be produced. A sample of the KWOC index to the thesaurus of *al-Jami'a* ( الجامعة) is shown in figure 7.4.

7.4.3 The PSI type and the Arabic language

The permutation technique has also never been tried on literature in the Arabic language. Obviously, this type of indexing involves more tasks than KWIC and KWOC types in the generation of keywords and the amount of sorting. On a computer produced printed page of a PSI type, there are two types of term. The primary (main) terms are always accompanied by some co-terms (secondary) which are indented as subentries and sequenced alphabetically under the primary terms which form the main access points to the index.

Although the description "permuted index" was used to describe some indexing products, the permutation technique of title keywords has never been applied in the Arabic language.
Figure 7.4: sample of the KWOC index to the al-Jami'a thesaurus
7.5 Producing a title-derivative printed index for Arabic articles

This section describes the procedures involved in the process of producing an Arabic title word printed index. It aims to reveal the limitations and the problems currently facing that process, starting with software selection and ending with problems during the data entry and the production stages.

7.5.1 Choosing the software

The absence of directories and tools which may disclose the functionality of the available Arabized and Arabic bibliographic computer software makes it difficult to track which of these is capable of producing printed title word indexes.

Apart from MINISIS, which apparently was used to produce the thesaurus of al-Jamī‘a KWOC section (see figure 7.4), and the system used by the Arab Centre for Medical Literature in Kuwait to produce a KWIC type printed list of medical terms (see figure 7.3), there are in the English literature of librarianship, some articles revealing that at least two of the previously mentioned Arabized software packages are capable of producing title word indexes.

Kanjilal (1992) mentioned that CDS/ISIS can be used for the production of a permuted keyword printed index. He developed a utility program using the Pascal programing language to generate an English permuted index. He also mentioned that permuted printed indexes can be generated for any repeatable fields like author, report number etc...

At the Liverpool University Library, the DOBIS system (1986) is used to produce an online English KWIC type index. It has been indicated that
before a record may be entered into the library catalogue, the DOBIS system forces the cataloguers to make a catalogue search by using the title search mode, resulting in a rotated title word index. The cataloguer keyboards a title word phrase and the system responds with a list of titles sequenced alphabetically in a KWIC format (see figure 7.5).

For the Arabized versions of these systems, it has not been indicated in the literature whether such capabilities have been considered during the Arabization process. Many institutions and libraries in the Arab world are currently using the Arabized versions of CDS/ISIS and DOBIS/LIBIS. However, no bibliographic product is reported to be using titles to generate title word printed indexes for articles using CDS/ISIS. Also, no answer was received from the Dean of Library and the supervisor of Arabizing DOBIS/LIBIS for King Fahd University of Petroleum & Minerals Library regarding this function. Muhammad Ashoor was asked through a letter whether the Arabized DOBIS system does allow cataloguers and users to make catalogue searches using the title search mode, and then shows the results in KWIC format. Unfortunately, no answer was received and no indication was given in the literature concerning the Arabized version of DOBIS/LIBIS.
### Catalog search

**Titles**

**Enter search term**

**Out of Africa**

**Catalog search**

<table>
<thead>
<tr>
<th>Title</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Out</em></td>
<td>a study of absenteeism among nurses</td>
</tr>
<tr>
<td><em>Which way</em></td>
<td><em>Out</em> a study of the Guyana-Venezuela boundary dispute</td>
</tr>
<tr>
<td><em>When the oil runs</em></td>
<td><em>Out</em> a symposium organised by the South Coast Sec</td>
</tr>
<tr>
<td><em>When the oil runs</em></td>
<td><em>Out</em> proceedings of the symposium organised by t</td>
</tr>
<tr>
<td><em>6</em> / <em>Oute regionale de l'</em></td>
<td>Outaouais</td>
</tr>
<tr>
<td><em>7</em> / <em>Habitat dans l'</em></td>
<td>Outaouais rural Quebecois</td>
</tr>
<tr>
<td><em>8</em> / <em>Into the Communal</em></td>
<td>Outbreak at Cawnpore and resolution of the Govern</td>
</tr>
<tr>
<td><em>9</em> / <em>The</em></td>
<td>outbreak of rebellion</td>
</tr>
<tr>
<td><em>10</em> / <em>The causes of the</em></td>
<td>outbreak of scurvy in the recent Arctic expedition</td>
</tr>
<tr>
<td><em>11</em> / <em>Into the recent</em></td>
<td>outbreak of smallpox</td>
</tr>
<tr>
<td><em>12</em> / <em>Propositions' and</em></td>
<td>outbreak of the Civil War</td>
</tr>
<tr>
<td><em>13</em> / <em>of James I to the</em></td>
<td>outbreak of the Civil War, 1603-1642</td>
</tr>
<tr>
<td><em>14</em> / <em>The</em></td>
<td>outbreak of the English Civil War</td>
</tr>
</tbody>
</table>

**Enter number or code**

**t new term**  **f forward**  **i new file**  **b backward**  **d detail**  **e end**
Although it is no longer difficult to tailor and modify existing Latin alphabet computer software that can generate automatic title keyword printed indexes and make them Arabic script compatible, it is observed that no interest is being shown in such type of packages. Reasons behind this unconcern may be that the issue of using the information content of Arabic article titles in producing printed indexes has not been properly considered yet. Certainly, showing that titles of Arabic articles can be used for producing printed indexes will lead to the consideration of such a facility while Arabizing existing software and writing programs for the Arabic language.

To fulfill the aim of this section and produce an Arabic title keyword printed index, two options are left:

i - To carry out the process manually, as was done by Watt in 1824 and by Andreas Crestadoro in 1856. They rotated by manual means the keywords in titles to produce the index to Watt’s Bibliotheca Britannica (see figure 3.1) and the book catalogue of the Manchester Free Library (see figure 3.2).

ii - To produce the index by means of computer, but with human intervention.

The first option is not feasible for this study because it is too time consuming, especially when the intention is that such indexes should enjoy quick production. Also because it will not assist in disclosing problems related to computer sorting and displaying of Arabic printed indexes. Therefore, the second option will be considered for the production of a title word printed index using the Arabized version 3.3 of a DataBase Mangement System called OMNIS.

The OMNIS database is the old-timer among Apple Macintosh relational databases, having been around since 1984. Its version 3.3, which was used for the index production, is the first major upgrade since OMNIS 3
became OMNIS 3 plus in June 1986.

Originally, OMNIS 3.3 was created by Blyth Software Inc. but was developed and made compatible with the Arabic language by TUSCO (The User Systems Company Limited) with the assistance of al-Hayat Information Centre which required the package to index the content of al-Hayat Arabic daily. Lehman (1988) mentioned that OMNIS version 3.3 has been on the market since September 1988. With regards to the Arabic version, correspondence between the OMNIS developer TUSCO and al-Hayat Publishing Company indicates that this version has been in use to handle Arabic script materials since August 1990. The aim of using OMNIS 3.3 was to enable al-Hayat to produce conventional printed subject and author indexes to its contents as well as to enable editors to consult the database regarding citations about published stories.

The Arabized input (data entry) window for the Arabic OMNIS 3.3 includes most fields needed for a newspaper database such as record number, publication date, issue number, page and column numbers, writer names, title and subject heading fields. As for report generation, the software can produce almost all kinds of reports required to correct and produce a conventional assigned subject heading and author index. In fact, the software gives options to create any of the following outputs:

i - A proof list of subject headings used for a certain index with or without their record numbers.

ii - A list of records used for a certain index with their numbers. Such records comprise all fields used for indexing.

iii - An author index arranged alphabetically according to author's surname.

iv - An alphabetical conventional assigned subject heading index with
The issue of evaluating this Arabized package is beyond the interest of this chapter. However, it was essential to find out whether it can, with its present functions, be used to carry out the production process of a title keyword index. What is already known from the manual for this package is that the Arabized OMNIS 3.3 in its present functions does not have the facility to generate title keyword printed indexes and is unable to suppress automatically prespecified stopwords from indexing positions. By this it is meant places where main entries or access points appear in a title keyword printed index.

Since this package is capable of producing subject indexes sorted according to indexer-assigned subject headings, it is evident that it is also capable of producing an Arabic KWOC type printed index with one condition. That is if the title keywords are humanly selected from the accompanying titles and then keyboarded in the fields originally provided for subject headings. Also, there is a possibility to include bibliographic information such as author names, the actual titles, source names as well as publication dates in the index entries.

Although the Arabic OMNIS 3.3 is not the relevant software for the automatic generation of Arabic title keyword indexes, it does meet certain requirements which make it relevant to be used for the production of a human aided title keyword index, needed in this research:

i - The software is Arabized. It handles the Arabic language and accepts all the Arabic characters.

ii - Ability of the software to produce Arabic subject printed indexes sorted alphabetically according to the ASMO 449 coding system specification.
iii - Permission to use the system in al-Hayat Information Centre was granted. Al-Hayat Information Centre is one of al-Hayat's departments situated in London.

iv - Availability of an Arabic system in al-Hayat. The use of the Arabic OMNIS 3.3 requires the use of an Arabic system compatible with the Apple computer. As an Arabic daily newspaper, al-Hayat depends on the Apple Arabic system in the production of the entire newspaper.

v - Ability of OMNIS 3.3 to produce an Arabic KWOC type printed index with indexer assistance.

7.5.2 The production process of the KWOC type index

To carry out this process, 110 Arabic article titles were randomly selected from twelve various disciplines. Article titles of ten of these disciplines were previously examined, in chapter five, against their humanly assigned subject headings found in al-Fihrist index. These are agriculture, economics, education, geography, history, law, philosophy, politics, psychology and sociology. In order to widen the subject coverage, two more disciplines were added. These are library and information science and linguistics.

The sample taken includes the first ten examined titles for each of the ten disciplines previously investigated, and five titles for each of the two added disciplines which were randomly chosen from al-Fihrist (الفهرست ), for linguistics and from the Message of the Library cumulative index ( رسالة المكتبة: الكشاف التراكمي 1965-1985 ) for library and information science. This cumulated index appeared in the Message of the Library ( رسالة المكتبة ), 21(4), 1986.

Since Arabic scientific articles are not covered in any Arabic indexing
tool, it was not possible for this study to consider titles from other Arabic scientific fields. Titles had to be taken from Arabic indexing tools so that their human assigned subject headings could be traced for the purpose of comparing between two types of printed indexes for the same Arabic articles. Section 7.6 deals with this comparison process.

In the data entry stage, a new data file called Arabic KWOC was opened on OMNIS 3.3, then the full titles of the selected articles and the keywords humanly extracted from the titles were keyboarded into the system in such a way that each title and its keywords were input into one record.

Since no modifications were made to this system for the sake of this process, the titles were keyboarded in the record fields specified for titles and the humanly extracted keywords were input in the fields originally specified for subject headings. Each of the extracted keywords was put in a separate field. The maximum number of fields allocated to subject headings in OMNIS 3.3 is 20. Fortunately, none of the article titles studied included more than 20 keywords, therefore all substantive words found in the examined sample were considered. The stopwords list used in this process was the same list which was used when the information content of Arabic titles was investigated in chapter four. This list is shown in appendix 4.2.

After the data entry stage and the manual suppression of the stopwords found in the selected titles had been carried out, OMNIS 3.3 was instructed to produce a subject index arranged alphabetically according to the keywords which had been already input to the system. It took the system about 5 minutes to produce two separate files containing 559 index entries comprising the sum of index entries for the entered keywords. 500 is the maximum number of entries that the Arabic OMNIS 3.3 can list in one file. So, the number of files produced was two. The first contained 500 index
entries while the second contained 59 index entries. Bibliographic details shown in each index entry were restricted to article titles and record numbers only.

In order to have the resulting two index files printed, they were imported to an al-Nashir al-Maktabi (الناشر المكتبي) Master document which consists of two columns. Al-Nasher al-Maktabi is an Arabized desktop publishing system used in al-Hayat with other relevant packages for the newspaper production. The printed index produced consists of 24 pages of A4 size. It includes 442 access points with 559 index entries. A sample page of the index is shown in figure 7.6.
العنف
النوعي الأوروبي : لحالة التضعف والعنف (14)
حجم
العلاقة الاقتصادية في حجم الزمن (29)
لا يوجد
أكتوبر : مادة الحلم (17)

العنف
النوعي الأوروبي

دراسة تحليلية في علم الإجتماع العسكري (18)

مصدر
الجوائز الإسلامية في القرن
السبع عشر ، بداية القرن السادس عشر (19)

العصر
من الدروس الدالى للعربية الفصى في العصر
الحديث (2)

العصر
حوت وسائل النضال المسلم الإسلامي الصليبي
في العصور الوسطى (41)

العقل
إبن رشد بين العقل والنقل (26)

العقول
حكم إجراء العقول بوسائل التنال الحديثة:
الحائط، البرقية، التلسك في هذ الشريعة
والقانون (37)

العقول
عمليات السلامة كمسلام علائي في خدمة الفرد
مع المشكولات السلكوية (42)

نمو
تطور علم التاريخ الإسلامي حتى نهاية القرن
العقاري (26)

ماهية علم الإجتماع وطبيعته وأهدافه (16)

علم الإجتماع بين التعددية والمركزة (26)

علم الإجتماع في الجزء هنغابيا (16)

دور الجماعة الأندية في التشكيلات القنالية:
دراسة تحليلية في علم الإجتماع العسكري (18)

علم
المعنى والمصادر في نمو المنهج التاريخي في
علم الثورية الإسلامية (17)

Figure 7.6 : sample of the Arabic KWOC type index produced for this study
7.5.3 Problems involved in the production process

The Arabic OMNIS version 3.3 which was used to produce the Arabic KWOC type index was not capable of fully conducting the production process without human intervention. Some of the emerged problems which demanded human intervention are related to the characteristics of the Arabic language (see chapter eight), and some others are caused by the use of the computer as a possible indexing tool that can handle and process Arabic script materials. The problems which were noticed and had to be humanly tackled during the production process are the following:

7.5.3.A The inseparable prepositions

The first hurdle that required human intervention was the Arabic inseparable prepositions that occurred at the beginning of some keywords. As a matter of fact and due to the Arabic language characteristics, inseparable prepositions such as the letters ل (l), ب (b), و (w) and ك (k) were treated as separable prepositions during the keyword selection and were stripped by manual means from every keyword in the sample. For instance the inseparable preposition و (و ) was manually stripped from the keyword والخليفة (والخليفة) translated in English as (and caliphs). The letter و (و ) was suppressed as a stopword and the keyword والخليفة (الخليفة) caliphs was considered the access point under the sorting letter خ (خ). If it had not been done this way, then the letter و (و ) would be the first sorting letter; something which brings confusion to the user who is looking for the keyword caliphs والخليفة (الخليفة) starting with the sorting letter خ (خ). The consideration of the Arabic inseparable prepositions and their effect on Arabic alphabetical sorting was discussed earlier in section 4.2.2.
It is relevant to mention here that in order to get better legibility, the principle of stripping inseparable prepositions from substantive words in the keyword fields was not applied when they occurred in the field provided for article titles. Article titles had to remain untouched with regard to the sequence of words and to the separable and inseparable prepositions that occurred in them. So, titles in the title fields came in the core of the produced title keyword index as they originally appeared in their articles. Otherwise, the titles in the index entries would seem disassembled and require more effort from the user to be read. As in the following disjointed title:

الوالدية المسؤولة و النمو السكولوجي ل الطفل

7.5.3.B The various types of the Arabic 'alif character

In addition to the human intervention stated above, a human also had to intervene in the determination of the alif character types used throughout the selected sample. The alif types intended here may also be called initial hamza types or hamzated 'alif character types. Within the course of this research the term alif will be used.

In the orthography of Modern Written Arabic, it is the practice among most Arab authors, editors and typists to substitute the bar 'alif ( ١ ) for the remaining forms of the available 'alif types occurring at the commencement of Arabic words in Arabic texts. For instance, Arab authors mostly write إيطاليا ( starting with bar 'alif ١ ) and not the correct form إيطاليا ( starting with 'alif which has its hamza down ١ ).

Typists who work for Arabic periodical publishers also adopt this practice. They generally typeset the Arabic letter bar 'alif ( ١ ) instead of the other types of Arabic 'alifs which are recognised by the computer.
that because they are used to it.

The three types of Arabic 'alif affected by this practice and by the substitution policy adopted by Arab authors and typists are the 'alif with madda above (ی), the 'alif with its hamza above (١), and finally the 'alif with its hamza down (٨). It is observed that the last two types of 'alifs are affected more by authors and typists.

Buckwalter (1992) mentioned the variation in the orthography of some Arabic characters, including letters and voweling marks. He specifically mentioned the orthography variation of the initial, medial and final hamza. He also mentioned the regional variation in the orthography of the final y (‘alif maqsurah, ی) being sometimes used instead of the dotted variant y (ی). Buckwalter (1992) considered these issues in connection with his study about such orthographical variations and their relevance to automatic spell checking in Arabic texts.

In the course of handling the sample of titles used in the process of producing the Arabic title keyword printed index for this study, there was no orthographical variation problem noticed with the medial and final hamza. Also there was no case of the undotted y (‘alif maqsurah, ی) being used instead of the dotted y (ی). This may be due to the fact that sample titles were taken from indexing sheets provided by the publisher of al-Fihrist index, and that indexers who filled in the indexing sheets handled other Arabic orthographic variations but overcame the issue of different types of 'alifs. Another possibility is that the problem of Arabic orthographic variation of some Arabic letters becomes more obvious when looking at larger texts such as the full texts of Arabic articles.
7.5.3.C Voweling marks

The Arabic voweling marks represent a special feature of the Arabic alphabet and are employed either to differentiate consonants or to represent vowels. They comprise the short vowels fathah, dammah, kasrah as well as the sukun, the shadda and the nunation signs fathatayn, dammatayn, kasratayn. The Arabic voweling marks do not exist in the main Arabic script but are written directly above or below the consonants they follow. In MAS (Modern Arabic Standard) it is quite rare to find such voweling marks in most printed materials.

In this study, among the 110 titles investigated only 3 had voweling marks in them. They appeared in the indexing sheets as the following:

1 - التعليم عن بعيد وإمكانية استخدامه في جامعة عربية مفتوحة
2 - البحرين: السنة الأولى من حملة محو الأمية
3 - خلوة القرآن التموذجية للأطفال

This finding does not necessarily mean that these three titles were the only titles with voweling marks before the indexers wrote them down on indexing sheets from their original articles. However, generally speaking the use of voweling marks in current periodical articles is rare and inconsistent. Buckwalter (1992) mentioned that voweling marks in Arabic language are rarely used in most genres of printed material, and if they are used, then they normally perform a disambiguating function. He showed that reading of an entire issue of the daily al-Hayat (الحياة) yielded the following exhaustive list illustrating typical usage:

رجع اليّ... عرض اليّ... نفذ أفكارك وأمضى... ولمّ لا؟... الكمّ الهائل... كمًا وكيفاً... مما يتمّ عن... وحسن تفهم... للحقائق... ارسل اليّ... الوزير العماني... لا أقرّ الشعر (ولا أقرّ الغلوس أيسرًا)...
Since the use of voweling marks in Modern Arabic Standard is rare and inconsistent, in this study voweling marks were omitted from the sorted keywords as well as from the index entries. The reason behind this option, the advantages and drawbacks are explained in the section about the Arabic voweling marks in chapter eight which discusses problems resulting when retrieving from Arabic title keyword printed indexes produced by computers.

7.5.3.D Indian numbers

The final factor which required human intervention during the production process of the pilot Arabic title keyword index is the existence of Indian numbers in the sample studied. Five titles of the sample had Indian numbers representing year dates but two of them each included two different dates defining certain periods. So the total number of year dates found in the titles was seven.

Although Indian and Arabic numbers representing year dates were counted as informative elements in the content of Arabic titles, it was observed that their use as access points would create confusion for users searching a printed index for information about certain dates or numbers. The Arabic version of OMNIS does not sort numbers properly when they are put in the fields provided for subject headings. For instance numbers such as 1919 and 1948 will appear between the numbers 1 and 2 in the sorted index. It seems that the system treats numbers as words and sorts them in accordance with the letter by letter filing system. Therefore these seven year dates were not added to the keyword fields during the data entry stage.

7.6 Comparison of two types of Arabic printed indexes

To find that the titles of Arabic periodical articles are representative
enough, and to find that the problems involved in the automatic generation of
the proposed Arabic title keyword printed index can be overcome, would not
be a complete answer to the proposal of depending on Arabic title keywords
in indexing the contents of Arabic articles. The end product, which is the
proposed printed index needs to be evaluated. Thus, the objective of the
following sections of this chapter, is to determine the effectiveness of the
Arabic KWOC printed index representing the title keyword indexing
technique. The performance of this index will be assessed in terms of
retrieval effectiveness against the predominant printed index to the Arabic
periodical articles al-Fihrist. (الفهرست), representing conventional indexing
techniques.

7.6.1 The evaluation process

Cleveland (1990, p.144) mentioned that an index can be evaluated either
as an individual product or in comparison with similar indexing products. The
aim of the first approach is to rate the printed index, in terms of the needs of
the end user, the fields covered, its stated purpose and objective, and its cost.
In the other approach the quality and cost of the product is compared with its
comparable indexes. But, to take the latter approach, " we must have made a
previous judgement about the other indexes ".

As far as the first approach is concerned, it is apparent that Arabic
literature researchers are in urgent need of indexing tools that can update their
awareness regarding their interests in various fields. As a matter of fact, the
proposal of depending on article titles in order to produce a cheap and quick
printed index with wider coverage arose from the fact that the available
Arabic indexing products are insufficient in terms of coverage, users' needs,
timeliness, and finally are not cost effective. Chapter two of this thesis
pointed out the deficiency of the current Arabic printed indexes.

To investigate practically the suitability and retrieval effectiveness of title keyword indexing techniques for Arabic literature, it was more appropriate for this thesis to follow the second approach of Cleveland. However, to conduct this process, there must be some kind of previous judgement about the comparable index. Thus, since the comparable al-Fihrist index has never been assessed, the following sections of this chapter are judging the retrieval effectiveness of the Arabic KWOC as well as al-Fihrist index, and then the quality of the first is compared with the latter. It is relevant to emphasize here, that there are in the literature about Arabic librarianship some studies and reviews which considered the al-Fihrist index, but none of them has employed end users in order to judge its quality (see section 2.3.2.B.2).

Over the years almost every possible indexing comparison among various types of printed indexes in the English language has been made. Aitchison, Lavelle and Hall (1970) and Aitchison and Hall (1973) tested the performance of four printed subject indexes. Keen and Wheatley (1978) carried out a test concerning the performance of nine printed subject index entry types. Coates (1976) while discussing the physical substrates in index evaluation, made reference to many studies concerned with printed index evaluation. Also, Madelung (1982) compared two types of printed subject indexes and made reference to various related studies. For more details about such evaluation studies see Foskett (1982, p.518) and Lancaster (1991, p.116) who devoted whole chapters to their reviews.

7.6.2 Producing an al-Fihrist type comparable index

To fulfill the requirement for a more objective comparison, an al-Fihrist
type printed index was produced using the assigned subject headings to the same 110 articles previously employed in the production of the KWOC type printed index. It was possible to collect these subject headings from the input sheets already prepared by al-Fihrist indexers to the contents of these 110 articles which were randomly selected from twelve various disciplines. An exception was the five library and information science articles that were taken from the Message of the Library cumulative index (1965-1985). Since the input sheets for these five articles were unobtainable, the article titles' reference codes found in the title index section of that index were used to extract all the indexer assigned subject headings allocated to each of them.

To produce the al-Fihrist type printed index for the sake of the comparison process to be carried out, a new data file called al-Fihrist was opened on OMNIS 3.3, then the article titles of the selected sample and their assigned subject headings were keyboarded into the system in such a way that each title and its subject headings were input into one record. Each of the assigned subject headings was put in a separate field and none of the articles selected was allocated more than 20 subject headings. Therefore, all subject headings found in the articles input sheets were considered.

After the data entry stage had been carried out, OMNIS 3.3 was instructed to produce a subject index arranged alphabetically according to the subject headings. It took the system about 4 minutes to produce one file containing 344 index entries. As in the KWOC type index, only the title and the record number were displayed with each index entry.

To have the index file printed, it was imported to an al-Nashir al-Maktabi Master document which consists of two columns. The al-Fihrist type printed index produced consists of 15 pages of
A4 size. It includes 284 unique subject headings with 344 index entries. A sample page of the index is shown in figure 7.7. It is relevant to mention here that all sorts of the hamzated 'alif which appeared in the assigned terms were also considered in the process of producing the al-Fihrist type index. This is unlike the case with the production of the real al-Fihrist, where indexers give the same sorting value for the 'alif with hamza above (١) and for the 'alif with hamza down (١), since the whole production process is carried out manually.
Figure 7.7: sample of the al-Fihrist type index produced for this study
7.7 Methodology of the comparison between al-Fihrist and the KWOC types

Index evaluation goes back to the early 1950s when ASTIA (Armed Services Technical Information Agency : now DDC - Defence Documentation Center) tried to establish whether the set of subject headings it had developed was more useful than the Uniterm indexing system devised by Taube. A large number of specialists have approached the issue of index evaluation in later periods. However, there has been no agreement on the methodologies employed to look at the performance and efficiency of printed indexes. In accordance with this fact, Cleveland (1990, p.143) stated that "no totally acceptable methods of evaluation have been agreed upon ", because "indexing is more of an art than a science, and, as such, it depends heavily on experienced judgement ". It seems that the matter of evaluating printed indexes is still a controversial issue, and it becomes more controversial when it refers to Arabic indexes. However, at the same time, it is essential that some method is adopted, that will assist in evaluating the available and the proposed indexing products.

As a matter of fact, the evaluation of Arabic printed indexes is a neglected field of librarianship and information studies and there are no previous studies concerned with index evaluation techniques similar to that established in the West. Therefore, the evaluation of the two mentioned Arabic indexing products should be seen as a first step in determining the relative effectiveness of two different types of printed indexes to the contents of Arabic periodicals: the al-Fihrist type index and the KWOC type index. The first was chosen because it is the most used printed index form in the Arab world, and the second because it is the indexing type proposed by this
To examine the retrieval effectiveness of these two printed indexes, nine queries comprising nine different topics and involving various levels of searching difficulty were prepared in accordance with the subjects covered by the printed indexes under examination. These queries were later distributed to 20 searchers who were randomly grouped in two categories (A) and (B). These searchers were randomly selected from the population of Arab researchers who were carrying out or had carried out higher degree studies. Searchers of the first category (A) were asked individually to retrieve the relevant articles to the queries using the al-Fihrist type index (see appendix 7.1). Searchers of the other category (B) were asked individually to find the relevant articles to the same nine queries using the KWOC type index (see appendix 7.2). This examination was set up so that 10 searchers would search the nine topics in the al-Fihrist index type and another 10 would search the same nine topics in the KWOC index type.

The queries representing the nine topic statements alongside their English translations were the followings:

1 - (look for articles on the Islamic concept of economy).
2 - (look for articles on combating illiteracy in Bahrain).
3 - (find articles on teaching in the Gulf region).
4 - (look for articles on the renaissance period).
5 - (look for articles on banks).
6 - (look for articles about children).
7 - (find articles about Judaism).
8 - (find articles about Ibn Rushd).
Each of these queries was composed to represent one case of the searching difficulties that a printed index user would normally encounter when conducting searches in real live situations. In addition to that, the use of different searching cases will reveal whether factors such as familiarity and experience in searching printed indexes would equally affect the search outcomes in the two types of printed indexes under examination, especially since no cross-references of any kind were used in both indexes. To verify that the employed cases of difficulty had covered the main cases that a searcher can face, they were inspected and agreed on by another professional. The nine various difficulties which were considered during the queries formulation are the following:

- Case one indicates a direct look-up of a subject heading (access point).
- Case two indicates the presence of more than one concept in the topic.
- Case three indicates the need to narrow the topic.
- Case four indicates the need to broaden the topic.
- Case five indicates the need to use synonymous terms.
- Case six indicates the use of singular terms to get more relevant articles to the topic, especially for the KWOC index type searches. In the Arabic language, the singular form of the term (‘atfal, أطفال) children is (tifl, طفل) child. Unlike the case in English language, this singular form term takes a different sorting position under a different starting letter.
- Case seven indicates the need to use related terms.
- Case eight indicates the need to look under a person's surname to find relevant articles.
- Case nine indicates the need to look under one word of a given country name to find relevant articles.
Having composed the topic statements and agreed on the cases of difficulty, the next stage was to designate two professional indexers jointly to identify the relevant articles to the nine queries. For this purpose, copies of the input sheets which belonged to the 110 articles used in the production of the two printed indexes under examination and a copy of the nine queries were passed to two indexers working in al-Fihrist index (الفهرست). This step seemed more appropriate than commissioning other indexers because indexers of al-Fihrist have got the expertise in indexing Arabic articles and were previously involved in the preparation of 105 of the total number of the used input sheets. The two indexers were also asked to suggest the length of time that an inexperienced user searching al-Fihrist type index would need to identify the relevant articles to all the queries; for this they suggested a maximum time of thirty minutes. The list of the input sheet numbers identified by the two indexers as relevant articles to the nine queries was later used to judge the retrievability of the KWOC and al-Fihrist type printed indexes. Out of the 110 articles, the al-Fihrist indexers identified 31 relevant articles to the nine queries, as follows: six relevant articles for query one, one relevant article for query two, six relevant articles for query three, one relevant article for query four, three relevant articles for query five, seven relevant articles for query six, one relevant article for query seven, one relevant article for query eight, and five relevant articles for query nine (see table: 7.5).

To gather the search results obtained by the searchers involved in this evaluation technique, the composed queries were put on a form which was duplicated and distributed to each of the selected searchers to fill out as they were conducting their search process. One searcher at a time, and in accordance with his category, was handed one of these forms with a copy of
the index to be searched and asked to start his search, bearing in mind that a time limit of 30 minutes was imposed on his search.

Apart from the queries, the forms included specific questions which asked if the searchers had ever used Arabic or non-Arabic printed indexes to locate items. It also included explanatory notes and definitions regarding the type of indexes under examination, their functions and the way they were sorted and produced. The forms distributed to searchers of the (A) category included information about the al-Fihrist type index, whereas forms distributed to searchers from the (B) category contained information about the KWOC type index. In addition to that, the forms asked the searchers to list the record numbers of the proposed relevant citations found under their specific queries. Since a time limit was imposed on searches, the length of time every search had taken was registered on the same form which belonged to that search.

Before starting his search, every searcher was requested to read the explanatory notes which described the product he was searching. He was reminded of the time limit allocated for his search and given clarification to any questions he posed regarding the aim of the process and the type of the index he was searching. He was also reminded to list the record numbers of the relevant index entries which he found under their specific queries and then left alone to carry out his search without any intervention until either the time limit was over or the search was accomplished.

7.8 Results of the comparison

Since the objective of this comparison process was to determine the effectiveness of the proposed Arabic KWOC index in terms of information retrieval, the search results obtained by its searchers were compared with the
results obtained by the searchers of the al-Fihrist type index.

To judge the effectiveness of the two examined indexes against the list of answers to the nine queries provided by the two indexers of al-Fihrist, the recall and precision measures were used. In fact, the use of these two measures was preferred because the criteria which were adopted in pursuing the comparison process do not contradict with the terms of using these statistical measures. Lancaster (1991, p.137) mentioned that effectiveness measures such as recall and precision are applicable to studies of retrieval from any type of database, whether in printed or machine readable form. He also mentioned that such measures can only be applied in cases where relevant items to the search topics are known in advance to be in the examined indexing tools.

Foskett (1982, p.19) and Cleveland (1990, p.149) defined the recall measure as the ratio of the relevant articles retrieved to the total number of relevant articles potentially available. Whereas the precision measure was defined as the ratio of the relevant articles retrieved to the total number of articles retrieved. These two measures are usually expressed as percentages by multiplying by 100. Buckland and Gey (1994) take these definitions further to say that for any given retrieved set "recall is a measure of effectiveness in retrieving (or selecting) performance and can be viewed as a measure of effectiveness in including relevant items in the retrieved set" and that precision is a "measure of purity in retrieval performance, a measure of effectiveness in excluding nonrelevant items from the retrieved set".

To find out how searchers of category (A) performed and how searchers of category (B) did in relation to the other category, each of the 20 searches executed was used to calculate recall and precision. First, the two measures were calculated on the query level, so that every query in a search was
counted. Second, to study the performance of a searcher for a whole search, the recall and precision was calculated on the searcher level. Finally, the recall and precision measures were calculated on the category level, so that the performances of the two categories could be compared.

7.8.1 The al-Fihrist type index (category A)

Using the two adopted measures to study the effectiveness of the al-Fihrist type index, the results displayed in table 7.1 were obtained. In addition to the average recall and precision for the (A) category, table 7.1 also shows the average recall and precision for every search carried out. It is reminded here that all searchers of this category depended on the al-Fihrist type index to locate the relevant articles to the nine queries posed to them.

According to the forms submitted by the searchers of this category, all of them have had experience with the use of printed indexes. However, some of them were experienced with using newspaper indexes or indexes to encyclopedias but not periodical indexes. The forms also showed that only two of the searchers had used the printed index to the Arabic periodical al-Fihrist (الفهرست). Four of the searchers of this category took the complete 30 minutes allowed for the search and the lowest time taken was 18 minutes. The average time that an experienced searcher took to conduct the requested search was 25.1 minutes. Table 7.2 shows the time taken by every searcher in this category to complete his search. It also gives the average time spent to answer the nine queries of the form.
<table>
<thead>
<tr>
<th>Searcher</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.68</td>
<td>0.75</td>
</tr>
<tr>
<td>2</td>
<td>0.65</td>
<td>0.90</td>
</tr>
<tr>
<td>3</td>
<td>0.58</td>
<td>0.82</td>
</tr>
<tr>
<td>4</td>
<td>0.68</td>
<td>0.91</td>
</tr>
<tr>
<td>5</td>
<td>0.48</td>
<td>0.88</td>
</tr>
<tr>
<td>6</td>
<td>0.29</td>
<td>1.00</td>
</tr>
<tr>
<td>7</td>
<td>0.68</td>
<td>0.88</td>
</tr>
<tr>
<td>8</td>
<td>0.77</td>
<td>0.80</td>
</tr>
<tr>
<td>9</td>
<td>0.77</td>
<td>0.80</td>
</tr>
<tr>
<td>10</td>
<td>0.77</td>
<td>0.96</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>0.64</strong></td>
<td><strong>0.87</strong></td>
</tr>
<tr>
<td><strong>S.D.</strong>*</td>
<td><strong>0.15</strong></td>
<td><strong>0.07</strong></td>
</tr>
</tbody>
</table>

Table 7.1: the average recall and precision of the (A) category.

* = standard deviation.
<table>
<thead>
<tr>
<th>Searcher</th>
<th>Time</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1*</td>
<td>20 min.</td>
<td>E</td>
</tr>
<tr>
<td>2</td>
<td>25 min.</td>
<td>E</td>
</tr>
<tr>
<td>3</td>
<td>30 min.</td>
<td>E</td>
</tr>
<tr>
<td>4</td>
<td>30 min.</td>
<td>E</td>
</tr>
<tr>
<td>5</td>
<td>25 min.</td>
<td>E</td>
</tr>
<tr>
<td>6</td>
<td>30 min.</td>
<td>E</td>
</tr>
<tr>
<td>7</td>
<td>18 min.</td>
<td>E</td>
</tr>
<tr>
<td>8</td>
<td>30 min.</td>
<td>E</td>
</tr>
<tr>
<td>9</td>
<td>21 min.</td>
<td>E</td>
</tr>
<tr>
<td>10*</td>
<td>22 min.</td>
<td>E</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>25.1 min.</strong></td>
<td><strong>E</strong></td>
</tr>
</tbody>
</table>

Table 7.2: the average time taken by searchers of the (A) category. E = experienced; * = Used al-Fihrist before.
7.8.2 The Arabic KWOC type index (category B)

Applying the same statistical measures on the search results of searchers of category (B) who used the KWOC type index produced the results shown in Table 7.3. This table displays the average recall and precision for each searcher as well as the average recall and precision for the search results provided by the searchers of the category (B).

According to the forms submitted by the searchers of this category, six of the ten searchers have had expertise in using printed indexes and four had never used an index before. Also, in this category, some of the experienced searchers have had their experience with using newspaper indexes or back of the book indexes. The forms also showed that only two of the six experienced searchers had used al-Fihrist Index (الفهرست). The length of time taken to complete the 10 searches of this category ranged between 15 and 25 minutes, and the average time was 20.5 minutes. Table 7.4 shows the time taken by every searcher in this category to complete his search as well as the average time taken to answer the nine queries of the search.

The results concerned with the length of time spent on every search can be analysed further to show that the average length of time taken by the experienced searchers of this category to carry out a search was 19.8 minutes and 21.5 minutes for the four inexperienced searchers.
<table>
<thead>
<tr>
<th>Searcher</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.48</td>
<td>0.94</td>
</tr>
<tr>
<td>2</td>
<td>0.45</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>4</td>
<td>0.61</td>
<td>0.95</td>
</tr>
<tr>
<td>5</td>
<td>0.52</td>
<td>0.94</td>
</tr>
<tr>
<td>6</td>
<td>0.55</td>
<td>1.00</td>
</tr>
<tr>
<td>7</td>
<td>0.68</td>
<td>1.00</td>
</tr>
<tr>
<td>8</td>
<td>0.74</td>
<td>0.96</td>
</tr>
<tr>
<td>9</td>
<td>0.74</td>
<td>0.96</td>
</tr>
<tr>
<td>10</td>
<td>0.48</td>
<td>1.00</td>
</tr>
<tr>
<td>Average</td>
<td>0.62</td>
<td>0.97</td>
</tr>
<tr>
<td>S.D.*</td>
<td>0.16</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 7.3 : the average recall and precision of the (B) category.

* = standard deviation.
<table>
<thead>
<tr>
<th>Searcher</th>
<th>Time</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18 min.</td>
<td>I</td>
</tr>
<tr>
<td>2</td>
<td>15 min.</td>
<td>E</td>
</tr>
<tr>
<td>3</td>
<td>20 min.</td>
<td>E</td>
</tr>
<tr>
<td>4</td>
<td>23 min.</td>
<td>I</td>
</tr>
<tr>
<td>5</td>
<td>25 min.</td>
<td>I</td>
</tr>
<tr>
<td>6</td>
<td>20 min.</td>
<td>E</td>
</tr>
<tr>
<td>7*</td>
<td>20 min.</td>
<td>E</td>
</tr>
<tr>
<td>8*</td>
<td>21 min.</td>
<td>E</td>
</tr>
<tr>
<td>9</td>
<td>23 min.</td>
<td>E</td>
</tr>
<tr>
<td>10</td>
<td>20 min.</td>
<td>I</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>20.5 min.</strong></td>
<td><strong>All</strong></td>
</tr>
</tbody>
</table>

Table 7.4: the average time taken by searchers of the (B) category. E = experienced; I = inexperienced; * = used al-Fihrist before.
7.9 Discussion

For a searcher seeking articles on a particular topic, the retrieval effectiveness of the indexing tool is very important. The previous section looked into this important aspect of index evaluation. It examined the retrievability of two types of Arabic printed indexes, the conventional type represented by the al-Fihrist type index and the unconventional type represented by the Arabic KWOC. The al-Fihrist type index was tested to act as a comparable index to the proposed KWOC index.

Attention may be drawn here to the fact that retrieval effectiveness constitutes only one aspect of index evaluation which requires the user participation in order to be judged. In the process of index evaluation, many aspects need to be tested such as timeliness, coverage, cost and other certain aspects which do not directly require the opinions of searchers, although these aspects are of concern to them. In fact some aspects like timeliness, coverage and cost were taken for granted as some of the main features of the title keyword indexing technique, and there was no need for such aspects to be investigated. After all, it was the deficiency of conventional indexing in these aspects that raised the need to devise unconventional techniques and led to the proposition of employing such techniques in indexing the contents of Arabic periodical articles. Chapter eight of this thesis looks at some of the aspects which affect the effectiveness of retrieval from Arabic printed indexes.

Lancaster (1991, p.74) defined good indexing in a very pragmatic way as "indexing that allows items to be retrieved from a database in searches in which they are useful responses and prevents them from being retrieved when they are not ". In accordance with this statement, an index can be judged
successful if it permits searchers to locate articles they want without having to look at many they do not want. Under this notion, the comparison of the two indexing types under examination was conducted to judge how searchers of the KWOC index type performed in relation to the other involved searchers who used the al-Fihrist type to retrieve the articles requested. To obtain a true picture of how the KWOC index performed, the recall and precision ratios were used together.

The results of this evaluation process did not contradict with the tendency which was noticed when tests of the performance of retrieval systems were first developed. Buckland and Gey (1994) stated that "an empirical tendency was noticed for recall and precision to be inversely related: one might have high recall or high precision, but not, it seemed, both at the same time. There appeared to be a trade-off, even though having high values of both at the same time is preferable. This trade-off is showed in the table of results displayed in the previous section of this chapter. The results showed that searchers of the (A) category achieved a recall ratio of 64% and a precision ratio of 87%, whereas searchers of the (B) category scored lower recall ratio of 62% but higher precision ratio of 97%. Using the T test of significance showed that the difference between the two categories is not significant at 1% level in terms of recall and time taken by searchers, but significant in terms of precision. The value of T was 2.86.

It was observed that searchers of the (B) category who used the KWOC type index to locate the relevant articles were able to achieve 97% of the recall ratio scored by searchers of the (A) category who relied on the al-Fihrist type index to retrieve the required articles. It should be borne in mind that 4 inexperienced searchers participated in the KWOC search and all searchers of the al-Fihrist type were experienced. The obtained results also
showed that searchers of category (A) who had used al-Fihrist index before, got an average recall of 73% against an average precision of 86% with 21 minutes as average length time. Whereas experienced searchers of category (B) who had used al-Fihrist index scored an average recall of 71% against an average precision of 98% with 20.8 minutes average time to conduct a search.

As far as individual performances were concerned, it was found that the highest recall was achieved by a searcher from the (B) category who scored a recall ratio of 94% with a precision ratio of 94% too. The highest recall scored by the searchers of the (A) category was 77% against a precision ratio of 96%. In fact this score was achieved by a very experienced searcher who has high qualifications in librarianship and familiarity with the use of al-Fihrist index, whereas the searcher from the (B) category was a research student in social sciences and had experience with using Index Medicus and Arabic newspaper and back of the books indexes but had never used al-Fihrist or any other Arabic periodical indexes.

In searching an indexing tool for relevant articles, the ideal is to achieve 100% recall and precision, and the desirable aim is to achieve very high percentages on both at the same time. However high recall (100% or close to it) is not always needed, since searchers commonly do not want all relevant articles. Most of the time they will be satisfied with what can provide them with recent articles or with some articles that can lead them to others. Buckland and Gey (1994) supported this notion and mentioned that searchers often prefer only one or a few relevant items to their queries.

Searchers who want to be absolutely sure of total recall can always ignore the printed indexing tool they are using and go straightaway to the bibliographic descriptions of the indexed articles, then read all of them. They
can examine the descriptions and check their full articles one by one until they have filtered out what they want. But, as Cleveland (1990, p.149) stated "this is the whole point of bibliographic control in general and indexing in particular - the filtering has already been done for the user. The index decreases the number of items users must examine, but since it is an artificial device, thrown between them and the material, there is the risk of not getting total recall ". Another possibility for searchers to achieve very high recall is to search sufficiently broadly to find all the useful articles, but this would probably lead to an intolerable precision. Lancaster (1991, p.4) mentioned that the larger the searched tool the less tolerable will be a low precision ". He went on to describe what the situation would be and said that "while a user might be willing to look at abstracts of say 57 items to find 6 useful ones, he may be much less willing to examine 570 abstracts for 60 useful ones. With very large databases, then, it becomes increasingly difficult to achieve an acceptable level of recall at a tolerable level of precision ". Therefore, searchers who insist on 100% recall would have trouble with precision, due to the inverse relationship between both measures.

In this evaluation process, some of the searchers in both categories achieved the ideal 100% recall ratio and the desirable high recall and precision in other times. But this was achieved on the one query level, which corresponded to one of the searching difficulty cases considered during the queries composition. However, when dealing with queries representing various cases of searching difficulty, none of the searchers achieved that recall ratio or complete precision ratio. The highest recall and precision ratios were achieved by a searcher of the (B) category who got 94% for both.

With regard to the performance of the two categories (A) and (B) towards each of the nine queries and their searching difficulty cases, it was
observed that in four of the nine queries searchers of the (B) category achieved better recall and achieved similar recall ratio in one of them. It was also observed that in eight of the queries, the same searchers achieved better precision and both categories had similar precision in one. The four queries in which searchers of the (B) category achieved better recall and precision ratios, correspond to difficulty cases number four, five, seven and eight. These searching difficulty cases are respectively concerned with the need to broaden the topic statement, the need to use synonymous terms, the need to use related terms, and finally the use of a person's surname to locate relevant articles. Whereas the query in which both categories had similar recall corresponds to the query number 2 which indicated the presence of more than one concept in the topic statement. Table 7.5 shows the performance of each category towards each of the nine queries.
### Table 7.5: The average recall and precision achieved by searchers of the categories (A) and (B) at the query level. * = relevant articles.

<table>
<thead>
<tr>
<th>Query</th>
<th>Recall</th>
<th>Precision</th>
<th>Recall</th>
<th>Precision</th>
<th>R.A.*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.72</td>
<td>0.93</td>
<td>0.70</td>
<td>0.99</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
<td>0.68</td>
<td>1.00</td>
<td>0.95</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0.50</td>
<td>0.69</td>
<td>0.39</td>
<td>0.88</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>0.40</td>
<td>0.28</td>
<td>1.00</td>
<td>1.00</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0.47</td>
<td>0.77</td>
<td>0.73</td>
<td>0.80</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>0.76</td>
<td>1.00</td>
<td>0.59</td>
<td>1.00</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>0.40</td>
<td>0.35</td>
<td>0.50</td>
<td>0.50</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0.90</td>
<td>0.95</td>
<td>1.00</td>
<td>1.00</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>0.60</td>
<td>0.90</td>
<td>0.56</td>
<td>1.00</td>
<td>5</td>
</tr>
</tbody>
</table>
7.10 Conclusion of the comparison process

Having found that Arabic article titles are representative enough to be used for title keyword indexing and for the production of title keyword printed indexes, it was crucial for the thesis to examine the retrieval effectiveness of such indexes, represented by the Arabic KWOC index produced for this purpose. Since the proposition was made of using a new indexing method to cover deficiencies on various levels found in covering Arabic articles of academic periodicals, it was also crucial to compare its retrievability with similar indexing tools currently used by searchers to locate relevant Arabic articles. Therefore the retrievability of the KWOC index was compared to the retrievability of the most used printed index in the Arab countries al-Fihrist (الفهرست). However, due to the absence of studies evaluating the retrieval effectiveness of al-Fihrist index, this chapter had to look at that and evaluate its retrievability through an al-Fihrist type index prepared for this purpose. To examine the obtained results, the recall and precision measures were used.

In accordance with what was stated by Cleveland (1990), Lancaster (1991), Buckland and Gey (1994) and various other authorities in this domain, and considering that the (B) category had some inexperienced searchers involved in the evaluation process, it can be said that searchers who used al-Fihrist type index did not perform significantly better than searchers who used the Arabic title keyword indexing in terms of retrievability. Although searchers of the al-Fihrist type index achieved better average recall, searchers of the KWOC type index achieved far better precision. Higher precision is always preferred, as it is an indirect measure of the cost of indexing since it reflects the number of articles that a searcher must look at in
order to identify the useful relevant articles to his search (Lancaster 1991, p.136).
Characteristics of the Arabic language and problems of retrieval from an Arabic title keyword printed index

This chapter is dedicated to revealing and discussing the problems which may be encountered by searchers when using an Arabic title keyword printed index for information retrieval and by publishers when producing it. It also sheds light on the characteristics of the Arabic language and its use in free text searching, since the proposed title keyword index is classified under the category of free text pre-coordinate indexing systems. The use of such indexes will be hampered by the problems of title keyword free text indexing discussed earlier, and in particular by the problems of Arabic language features. Therefore, it is relevant to start this chapter by exploring the Arabic language and its characteristics, then discussing the main aspects and problems of retrieving from the proposed indexing tool.
8.1 Introduction

Before going into the details of this chapter, it is relevant to confirm that the sample Arabic titles used in every examination process in this thesis represent the literary Arabic language as employed by contemporary Arab authors, and taken from current Arabic periodicals which use Modern Arabic Writing as a medium of communication.

By literary Arabic is meant the uniform variety of Arabic language used by academic and educated Arabs for a variety of linguistic functions. This type of language is used all over the Arabic speaking countries as the common medium of written communication in books, periodical articles, business and personal letters and memoranda. It is also used as medium of oral communication in formal speeches and in radio and television broadcasts as well as on occasions where some degree of formality and solemnity is required.

Literary Arabic represents a continuation of classical Arabic and differs from colloquial Arabic dialects (which differ considerably among themselves and from country to country and area to area in the same country). The colloquial dialects are used for every day oral communication by all people of the dialect area. Literary Arabic is described in the literature in many different terms. Some authors call it Standard Arabic, other some call it Modern Standard Arabic and others call it Modern Written Arabic. These various terms are used to distinguish between the standard varieties of Arabic of the present time and the classical Arabic language. Mehdi (1986) stated that "the difference between the two lies basically in the lexicon and style and to a much lesser extent in grammar " and that both literary and classic Arabic " differ from the colloquial version in the lexicon, style, phonology, syntax,
and sociolinguistic function ".

8.2 Arabic alphabet

The Arabic alphabet taught to students at their primary schools consists of 28 characters, but this is apart from the lam-'alif ( \( \text{ل} \) ) which is a combination of two characters 'alif ( \( \text{ا} \) ) and l ( \( \text{ل} \) ) that take one special shape whenever the Alif follows the Lam. In some stages of primary education, this letter is to be counted as one character before the y ( \( \text{ي} \) ) which is the last character in the Arabic alphabet. Arabic is written from right to left with no capital letters, and most letters in the Arabic word are joined together. Abboud (1972) divides the letters into two major categories: six non-connectors which always retain their shape but can be connected only to a preceding letter, and twenty two connectors which can be joined on both sides and in the process take different shapes depending on their context in a word. The position can be in the beginning of the word like the letter b ( \( \text{ب} \) ), or in the middle of the word ( \( \text{ب} \)), or at the end of the word ( \( \text{ب} \)). The letter can also be written separately not connected to another letter in the same word ( \( \text{ب} \)). When counting the Arabic characters, Aman (1984) added to them three others and mentioned that the Arabic alphabet consists of 31 characters. To these he added the hamza ' ( \( \text{ه} \) ) which appears as a separate character in the written language, but is rarely used alone, t marbutah ( \( \text{ت} \) ) which is one of the possible shapes for the letter t ( \( \text{ت} \) ) when it appears as a stand-alone terminal character, and 'alif maqsurah ( \( \text{ى} \) ). However, to use computers to handle Arabic materials, ASMO (the Arab Organization for Standardization and Metrology) assigned 36 characters for the Arabic alphabet including the various types of the hamzated 'alif ( \( \text{ٰ} \) ) and hamza ' ( \( \text{ٰ} \)) as well as the t marbutah ( \( \text{ت} \) ) and the 'alif maqsurah ( \( \text{ى} \) ). The
Arabic characters and their various shapes within a word are shown in appendix 8.1 which is taken from Wright (1933, p.1).

Besides the alphabet character set, Arabic has a certain number of voweling marks including the vowels which are written in the form of over and underscores of the consonants they follow. In the Arabic language there are at least three signs denoting these sort of vowels. These are:

a - The fathah which is a small stroke written above a consonant \( \text{ـ} \), as in the letter \( \text{ب} \) (ba).

b - The dammah, a small character similar to the Arabic letter (و) above a consonant \( \text{ـ} \), (like the English comma), as in the letter \( \text{ب} \) (bu).

c - The kasrah, a small stroke under a consonant \( \text{ـ} \), as in the letter \( \text{ب} \) (bi).

Semaan (1968) mentioned that 'Ibn Jinni, an early Arab grammarian, said that the Arabic vowels are parts of the letters of prolongation, the so-called soft letters, namely the 'alif (ا), the w (و) and the y (ي) and just as these letters are three, so are the vowels three in number, namely the fathah (the open), the dammah (the lip-rounded), and the kasrah (palatal). The prolongation letters are sometimes referred to as long vowels and the voweling marks are known as short vowels. When counting the letters, some authors add another two signs. Aman (1984) mentioned the shaddah and the sukun. Wright (1933, p.13) said that the first is marked with the sign \( \text{ـ} \) which is used for doubling a consonant, whereas the latter is marked with the sign \( \text{ـ} \) which is written over the final consonant of all shut syllables, and serves, when another syllable follows, to separate the two.
8.3 Arabic language features

Apart from the features revealed in the previous two sections, the Arabic language structure is different from that of English or other Indo-European languages. The most characteristic feature of it is that the great majority of its words are built up from roots, each of which consists of three consonants or radicals. Hegazi and al-Sharkawi (1985) claimed that words in common use in Arabic language are based on about 1200 different roots. By using these roots or radicals as a base and adding prefixes, infixes and suffixes, according to certain patterns, the actual words are produced. Yahia (1989) cited 120 different forms of nouns resulting from adding affixes to the basic naked noun. It is relevant to mention here that in the Arabic language, adjectives, adverbs, and pronouns are treated as nouns.

Another feature of Arabic language is mentioned by Mehdi (1986) who said that "Arabic, like Latin, is a synthetic, or inflectional, language rather than a language like English which is predominantly analytic ", and that the "the syntactical relationship of nouns is indicated by case endings and verbs are inflected by means of prefixes, infixes, and suffixes to indicate the various persons, numbers, genders, derived forms, moods, and tenses, in contrast to English where, for example, a separate word (noun or pronoun) is required to indicate the person ".

8.4 Retrieval problems

Apart from the general retrieval limitations of a pre-coordinate free text retrieval tool, the proposed title keyword printed index suffers from other particular retrieval problems which stem from the characteristics of the Arabic language. The fact that Arabic is a highly inflected language leads to
the assumption that this might have an effect on the efficiency of the language in the free text information retrieval technique. However, in this thesis it is assumed that infixes and suffixes will not considerably disrupt the alphabetical sequence of a printed index. Infixes occurring in the middle of a word and suffixes which occur at the end of it, will rarely affect the alphabetical order of a printed index.

Generally the retrieval problems which were found when producing a KWOC title keyword index for this study can be grouped in three categories. The first category includes problems which can be tackled during the input stage. The problems of the second category can be partially overcome with the development of sophisticated software specially to deal with Arabic free text indexing. The third category includes problems which would be diminished when Arab authors start to compose more relevant article titles for the purpose of derived indexing.

The following sections of this chapter will discuss the problems which are included in the the first and the second categories. Problems of the third category will be discussed in chapter nine as they constitute a part of the recommendations to authors.

8.4.1 Retrieval problems handled at the input stage

The Arabic language characteristics which come under this category will not cause any problem during the retrieval process, if they are properly tackled prior to the production stage. Since the capabilities of computer software which is currently used to index Arabic materials cannot tackle them automatically, the intervention of the human indexer is inevitable. Otherwise they will add to the problems which may be encountered by searchers. These characteristics are the following:

296
8.4.1. A Usage of voweling marks

As mentioned earlier in section 7.5.3.C, the Arabic voweling marks represent a special feature of the Arabic alphabet. They include all sorts of vowel and orthographic signs such as the fathah, dammah, and kasrah as well as the sukun, the shaddah and the various types of tanwin which is marked by doubling final vowels at the end of words. Tanwin is also known by orientalists and Arabists as nunation. The Arabic voweling marks are used to vocalize the words and may be referred to as vocalization symbols.

Originally, Arabic writing did not have voweling marks but in later stages their role was recognized and emphasized by early Arab grammarians. Kasem (1978, p.179) quoted that according to al-Nawawi, adding vocalization symbols to the scripture is "desirable because it is a guard against corrupt pronunciation and alteration". This desire of early Arab grammarians is currently disregarded by contemporary Arab writers. 'Ali (1994, p.356) stated that the neglect of using voweling marks has become a constant habit in reading and writing Arabic texts. Moreover, he thinks that vocalization will only be used as a medium for educating Arab children in the principles of speaking and writing the Arabic language. Although vocalization symbols are essential for the semantics of Arabic sentences and words, their current usage is optional in writing and they are not customary in commercial type applications. They are merely used when it is not possible to guess the correct word from the context. Currently, the use of voweling marks in Arabic contemporary writing including periodical articles is rare and the correct reading of Arabic texts is dependent upon the reader's intuition and linguistic or grammatical knowledge as well as the surrounding contexts of the ambiguous words.
Since Arab authors are inconsistent in providing voweling marks and rarely use them in their article titles, producers of Arabic title keyword indexes cannot commit themselves to the use of voweling marks with articles selected for indexing. Otherwise, grammarians will be involved in the preparation and will add costs to the production process. It is easier and cost effective to ignore all types of voweling marks either by means of computer when software is ready for this or manually during the data entry stage.

As a matter of fact the absence of voweling marks from index entries will not cause deficiency in retrieving, and Arabic periodical indexers never add them to subject headings or keywords assigned for articles. To avoid homographs, they provide a sort of scope note beside ambiguous access points. Figures 8.1 and 8.2 respectively show examples from al-Fihrist and the Islamic Index.

After all, the consideration of factors such as the surrounding contexts of access points provided by most title keyword indexes and searchers' presumed grammatical knowledge will reduce the number of problems from homographs dramatically. Kasem (1978, p.192) revealed that ignoring vocalization symbols in 117 article titles resulted in creating only one homograph in his pilot experiment and the number increased to eight artificial homographs when he increased the sample studied to 434 article titles for the main experiment.
Figure 8.1: Sample of al-Fihrist index showing the use of scope notes beside ambiguous subject headings.
Figure 8.2: sample of the Islamic Index showing the use of scope notes beside ambiguous keywords
8.4.1.B Inseparable prepositions

One of the main extra features needed in software for generating Arabic printed indexes is the capacity for automatically stripping inseparable Arabic prepositions from their accompanying keywords. The principle is to input the selected titles to the system once, and then it is the job of the system to carry out the extraction of substantive words and the suppression of stopwords.

It is a serious challenge to software developers experienced in the Arabic language to develop such software, especially since some of the inseparable prepositions are sometimes part of the original keywords. For instance, the keyword وطنية (wataniyyah, patriotism) starts with the letter w (و) which forms a real part of it. However, in some cases this letter precedes a keyword as an inseparable preposition like the case mentioned in section 7.5.3.A.

Only the emergence of software which can execute the stripping of the inseparable prepositions automatically will make the dependence on human intervention to conduct such a process unnecessary. As mentioned in section 7.5.3.A all existing inseparable prepositions will then be treated as separable ones and suppressed from the indexing position by means of computer.

8.4.1.C Variation in the orthography of some letters

Human intervention to determine the correct formal written shapes of the Arabic alif during the data entry process was necessary for the production of an Arabic alphabetical printed index sorted by means of computer for two main reasons:

i - The national and international standard coding systems designed for Arabic language character sets processed by computer assign to the Arabic
letter 'alif more than one coding value (Clews, 1988b). Character sets show which bit combinations (zeros and ones) function to represent specific characters or letters in computer systems. Clews (1994) mentioned that this coding "becomes a standard when all users agree on the same such usage in their own systems. If it is agreed within the industry only, it is an industry standard: if it is agreed by national or international standards organizations it is a national or international standard character set". For instance, ASMO (the Arab Organization for Standardization and Metrology) has issued a 7-bit character set standard referred to as ASMO 449. It was made international with the provision of an identical character set issued by ISO (the International Organization for Standardization 9036). Each Arabic letter in this character set has been assigned a unique bit sequence, including the different types of hamzated 'alif and the bar 'alif. For example, the 'alif with madda (١) is represented by the bit sequence 1000010; the 'alif with hamza above (١) is represented by the bit sequence 1000011; the 'alif with hamza down (١) is represented by the bit sequence 1000101; and the bar 'alif (٠) is represented by the bit sequence 1000111.

This assignment is of crucial importance, especially when it comes to the use of a computer to produce a printed index, because data fields will be sorted (in our case the keyword fields) according to the bit sequence. So keywords beginning with 'alif with the hamza above (١) such as the keyword أميّة ('Ummiyiah, literacy) will precede keywords that begin with 'alif with the hamza down (١) such as the keyword إسلام ('Islam), because the value of the bit sequence 1000101 is less than the value of the bit sequence 1000011. Figure 8.3 shows these four types of Arabic 'alif as the following:

a - 'Alif with madda above (١ corresponds to hexadecimal C2 in the figure).
b- 'Alif with hamza above (ı corresponds to hexadecimal C3 in the figure).

c - 'Alif with hamza down (j corresponds to hexadecimal C5 in the figure).

d - Bar 'alif (ı corresponds to hexadecimal C7 in the figure).

It is relevant to mention here that the latest 8-bit character set ASMO 708, mentioned earlier in this thesis, also assigns four different coding values for the Arabic alif, and was made international with the provision of an identical character set issued by ISO (the International Organization for Standardization 8859).

ii - The bar 'alif (ı) could not be adopted as a replacement for the other 'alif types, because it is an integral part of the definite article 'al (ال) called in Arabic أدات التعرف (the instrument of definition). It plays the role played by the English article (the) when it occurs before keywords, and is used in Arabic as much as the article (the) is used in English. The obvious difference between the English article (the) and the Arabic article 'al (ال), is that the latter is always written in conjunction with its following word. For instance, the English term (the book) consisting of two words, matches the term (الكتاب) consisting of one word in Arabic. Since in Arabic, the definite article 'al (ال) and every following keyword are written continuously, Arab indexers and cataloguers believed that to ease searching, this definite article should be ignored in the filing processes. Therefore, they are accustomed not to count it when preparing manual printed alphabetical bibliographic tools. The Arabic printed indexes represent one kind of these tools.

Other variations which have to be considered are the final y ('alif maqsurah, ى) being sometimes used by some Arab authors, especially the
Egyptians, instead of the dotted variant y (١) and the ta marbutah (٩) being sometimes used instead of the letter h (٨) when it comes at the end of the word. However, these two characters do not disrupt the sorting order of a printed index since they can not be mistaken or wrongly spelled when they occur at the start of a keyword.

To make retrieval from a printed index easier for searchers who are accustomed not to differentiate between the 'alif with hamza above 'a (١) and the 'alif with hamza down 'i (٩) during a search process, it is suggested that software developers consider giving these two letters the same sorting value in programs used for the production of Arabic printed indexes. Such programs will then sort keywords starting with 'a (١) and 'i (٩) in the same place. Users of CDS/ISIS have the option of giving two different letters the same sorting value. For instance, the upper case and the lower case forms of the English character A can appear in the same sorting category.
Figure 8.3: Arabic MS-DOS code page 708 (compatible with ASMO 449 and 708) showing the various types of 'alif.
8.4.2 Retrieval problems requiring sophisticated software

Unlike the previous category which included problems particularly related to the characteristics of the Arabic language, this category includes some problems which are common in producing title keyword indexes by means of computer no matter what type of language used.

8.4.2.A The definite article

There is no indefinite article in Arabic. The definite article for all cases, numbers and genders is 'al (ال), which is prefixed to the word it defines.

Van de Vate (1986) mentioned that the parts of an Arabic name often used as the access point frequently start with the definite article 'al (ال), and stated that "though in Arabic script the article and the name are written continuously, the article was in the past always ignored for filing purposes in both Arabic script and transliterated catalogues ".

To apply the convention of ignoring the definite article 'al (ال) with the use of a computer as a sorting tool is not as simple as applying it manually. It is not possible yet for the available Arabic indexing software to distinguish between the bar 'alif (١) when it is a part of a definite article 'al (ال) and when it is a hamzated 'alif before a keyword starting with the letter ل (ل) such as the keyword (الل). This keyword, if written with bar 'alif will appear as (الل), then the software, which is already instructed to suppress 'al (ال) will sort it under the letter m (م), that is the third letter in that keyword.

Currently, most DBMS (DataBase Management Systems) and indexing software compatible with Arabic script can handle the problem of the definite article easily, provided that the other types of the Arabic hamzated 'alif (١١١) are correctly used. The Arabic OMNIS version 3.3 is instructed to suppress
automatically the definite article 'al (اَل) from the sorting process, and the alphabetical sorting of printed indexes produced using the Arabic OMNIS version 3.3 is made according to the third letter in each access point beginning with the definite article 'al (اَل). The problem of the Arabic definite article 'al (اَل) will not arise if the convention of ignoring it in software which may be used for the production of Arabic title keyword indexing is considered.

8.4.2.B Existence of synonymous keywords

A main deficiency of title keyword indexing is that similar keywords are scattered throughout the index main entries, whereas in controlled indexing all equivalent keywords are gathered together under one keyword. Synonyms are words which have similar or nearly similar meanings in the same language but differ in spelling.

The discussion of the issue of synonym occurrences in an Arabic derived title keyword printed index, is in fact a discussion of synonymous title keywords as they are used by authors and searchers in the contemporary Arabic language. Kasem (1978, p.181) stated that the Arabic language is rich in synonyms. This is in harmony with other studies which approached the Arabic language and its characteristics. Thus to produce an Arabic title keyword printed index, there has to be some kind of control to alleviate the problem of similar concepts being scattered through the index main entries. The use of see and see also cross-references is a solution for alleviating the degree of subject scatter, arising from the use of synonymous keywords.

An early attempt at using cross-references in English language title keyword printed indexes was done in 1960 by DTIE (the Division of Technical Information Extension) of the United States Energy Commission.
during the preparation of a bibliography about radiobiology which was entitled *The Effects of Radiation and Radioisotopes on the Life Processes*. Voress (1965) mentioned that during the preparation of the preliminary rotated title keywords index of that bibliography, a preliminary deficiency resulted from the lack of standard language used in the selected title keywords which were rotated and caused similar keywords to be scattered throughout the index. In an attempt to ease this problem, an extensive cross-reference listing for *see* and *see also* was prepared using both the frequency keyword list and the preliminary rotated index. To produce the index DTIE used an IBM-1401 package which provided the option to input title keywords with humanly inserted cross-references. Currently, it is believed that the option of inserting cross-references automatically is possible with the huge variety of options such as Check, Edit, and Replace which are provided by indexing software.

As far as Arabic printed indexes are concerned, none of them depends on software options to insert *see* and *see also* references automatically. However, the idea seems possible with the provision of *al-Jami'a* thesaurus in machine-readable form compatible with the MINISIS system. The way that Arabic compatible indexing software may be made or developed to provide such options is for further studies to investigate after Arabic title-derivative indexing is widely tested.

It is relevant to mention here that since title keyword indexing depends on a relatively short part of an article in comparison with abstracts and full texts, the probability of synonyms occurring within one title is much lower than within one abstract or one text. Throughout the KWOC index produced for this study, searchers scored 73% recall and 80% precision for the query number 5 which represented the case of synonyms.
8.4.2.C Existence of plural and singular forms for the same keywords

The problem of keywords scattering and the various sorting orders that keywords having the same word root may take constitutes another drawback for title keyword indexing. However, this problem is not unique for Arabic language printed indexes. Title keyword printed indexes in Western languages also face this problem, and there have been some attempts to alleviate it. To tackle the problem, Voress (1965) also described in his study which handled the problem of synonyms, the step taken to bring together singular and plural forms of a keyword. He stated that a list of singular and plural forms of keywords and keyword variants was prepared so that keywords so related could be filed following one another by the computer program regardless of their normal alphabetical position.

In the Arabic language, keyword variants such as طفّل (tifl, child) and أطفال (‘atfal, children) resulting from the broken plural require such an option to exist in any program used for the production of a title keyword index. The broken plural in Arabic is constructed according to some common patterns in which the singular form needs to be completely changed to generate its plural form. Other plural forms such as the sound masculine and the sound feminine do not cause major problems in a printed index since the form of the singular remains the same as in the plural for the starting letters of a keyword. Throughout the KWOC index produced for this study, searchers scored 59% recall and 100% precision for the query number 6 which included the case of one keyword in both singular and plural forms inflected according to the Arabic broken plural and taking two different indexing positions in the index.
Another problem in retrieving from a title keyword printed index is the inconsistency in Arabizing foreign terms used in the contexts of Arabic articles. Arabized terms include proper names for individuals and places, as well as subject descriptors. It is believed that this problem faces searchers of Arabic indexing tools more than any other languages. Arab authors do not stick to Arabization rules of foreign terminology provided by the Arab League on one hand and on the other they rarely use Arabic equivalents to foreign terminologies in their texts. For instance authors in the library and information field use the word ببلیوغرافیا (bibliyughrafya) which is an Arabization for the word bibliography and do not use the equivalent term الورائة (al-wiraqah) which is used by early Arab bibliographers. In addition to that there is always the problem of being inconsistent in writing the Arabized terms. For instance in some texts the word ببلیوغرافیا (bibliyughrafya) with the letter ج (j) is being used instead of the letter غ (gh) and sometimes another letter is added and the word becomes ببلیوغرافیا (bybliyughrafya, third form with the addition of the dotted variant ي 'y' right after the first letter of the word).

In fact it is rather difficult to survey the Arabized terms used in Arabic texts and no such comprehensive list is available, but it can be said that inconsistency in Arabizing foreign terms is a major problem in information retrieval. There are always new terms and new names to be added to this survey. Another problem is the inconsistency in assigning equivalent Arabic terms to foreign terms. A recent term Privatization has been assigned at least four Arabic terms which are currently being used in the field of economics. These are: (ال-'istikhsas، الاستخصاص؛ الخصوصية،الخصخصة؛ al-takhsis، al-khaskhasah، al-khusuwwsiyyah).
It seems that the only solution to the problems of synonyms and keyword variations is the use of a machine-readable thesaurus intended for indexing or a word list linked to software capable of carrying out the functions of checking input for correctness of spelling and indication of synonymous and related keywords through the automatic generation of see also references which are needed in a particular issue of a printed index. The use of such a tool is virtually a necessity and the only effective means to avoid deficiencies arising from the existence of synonymous and related keywords and from the keyword spelling variations which appear in Arabic article titles.

The creation of a machine-readable thesaurus or an automated word list intended for producing pre-coordinated title keyword printed indexes is a whole project in itself.

8.5 Conclusion

Users of Arabic title keyword printed indexes will be hampered by some retrieval problems. These problems can be alleviated by the introduction of Arabic script compatible software packages able to perform sophisticated functions. The absence of such software packages requires the users to depend on their knowledge of the terminology used in the searched topics as well as their understanding of the Arabic language structure.

As a matter of fact, these drawbacks are not unique to the Arabic language. Users of title keyword printed indexes produced in other languages also face retrieval problems. However, the actual production and usage of such printed indexes as well as the extensive research conducted on the retrieval problems have assisted in minimising such drawbacks.

As far as Arabic title keyword indexes are concerned, it is believed there
is an interrelation between their production and the alleviation of retrieval problems resulting from their use. To suggest solutions to retrieval problems, title keyword printed indexes have to be produced and tried. Solutions to problems can then be sought and the use of titles in these services should affect the practice of authors and publishers. Chapter nine of this thesis is dedicated to the conclusion of this research with regard to the possibility of producing Arabic title keyword printed indexes and their use as bibliographic retrieval tools for Arabic periodical articles.
Chapter nine

Conclusions and recommendations

9.1 Background to the study

The complexity of conventional indexing requirements, the deficiency and ineffectiveness of the current indexing systems for the contents of academic Arabic periodical articles, the scarcity of Arab indexers with the appropriate indexing expertise and the growing concern regarding time, coverage and cost factors in the preparation of the current Arabic printed indexes have brought about the demand for new indexing methods to meet the increasing information needs of researchers using Arabic literature. The computer generated title keyword index was introduced for Western publishers and researchers as a fast and cheap means of indexing the growing volume of literature. This technique reduces or eliminates the human indexer effort and permits quicker index production at a lower cost and wider coverage. The objective of this thesis was to study whether titles of Arabic academic periodical articles can be used as basis for the production of Arabic title-derivative printed indexes by means of computer.
9.2 Procedures followed in the thesis

To provide a complete answer to this question, the thesis had to examine all the related issues to the process of using computers to produce Arabic printed indexes. The main issues involved were the Arabic titles and their informativeness as well as their representation, the computer hardware and software, the user reaction to the proposed printed index as well as the characteristics of the Arabic language regarding the retrieval from a title keyword printed index.

a - All available types of Arabic printed indexes which cover the contents of Arabic periodical articles were examined.

b - Almost all the available types of title keyword indexing techniques which cover the contents of English periodical articles were discussed, as well as the functions of Arabic article titles in the current indexing tools.

c - Approximately 1159 Arabic titles were employed in the processes of examining the information content of Arabic titles and their representation of the accompanying articles.

d - The retrievability of the title keyword indexing, represented by the Arabic KWOC index produced for this purpose by means of computer was compared with the retrievability of the most used printed index in the Arab countries al-Fihrist (الفهرست), and the retrieval problems of using Arabic title keyword printed index were discussed. Twenty Arab users were involved in this process.

9.3 Findings of the thesis

Findings in relation to the topic of this thesis may be summarised as follows:
9.3.1 Arabic periodical indexes

It is observed that conventional printed indexes for Arabic periodical articles generally suffer from non-currency and unsatisfactory coverage of periodicals published in the Arabic language. Many of them had started as indexes for current Arabic periodicals before they announced themselves to be retrospective or turned to handling backdated periodical articles. Reasons behind this may be the dearth of Arab indexers with appropriate indexing expertise and the conventional indexing processes which are being used.

9.3.2 Titles of Arabic articles

The absence of title keyword indexing techniques from the Arab world does not necessarily mean that titles of Arabic periodical articles do not have other functions in the current Arabic indexing tools. It is shown in the course of this thesis that titles do have other functions to play alongside the other bibliographical information of Arabic periodical articles. This absence also does not necessarily mean that titles of Arabic periodical articles are inadequate to be used for such indexing techniques.

9.3.3 Informativeness of Arabic titles

The investigation into the informativeness of titles was carried out by using a straightforward technique which involved the manual counting of all words and substantive words found in an article title. This technique has been used by several authors for English titles (see section 4.1.1). Although indexing by each keyword in periodical article titles was not properly considered or applied before, Arabic titles appear to be as informative as English titles in the 16 scientific and non-scientific subjects examined.
Several studies have examined the information content of titles of English periodical articles and found them to be sufficiently informative to be used for indexing. Since titles of Arabic periodical articles are found to be as informative as English titles, then they are sufficiently informative.

9.3.4 Representativeness of Arabic titles

The representation of titles of Arabic periodical articles was investigated by employing two techniques. The first involved the matching of title keywords and their corresponding subject headings assigned by human indexers. It was found that 46.7% of the subject headings humanly assigned to the examined articles had their access points similar to keywords in the titles. This technique was applied to titles of English periodical articles before by several authors (see sections 5.2 and 5.3). But it was the first time to be employed for studying the representation of Arabic titles. Also, it revealed that al-Fihrist is not really an indepth indexing tool. The lowest mean number of title keywords, represented by sociology, is higher than the highest average number of subject headings found in politics. In all examined disciplines, the numbers of keywords found in titles were more than their humanly allocated subject headings. In terms of exhaustivity, an Arabic title keyword index would be more exhaustive.

The second technique involves the examination of the occurrences of title keywords in the corresponding topic sentences of the selected articles. This process found that a mean ranging from 70% to 88% of the investigated title keywords occurred in the topic sentences. The high percentage of similarity found between the Arabic title keywords and the keywords in topic sentences indicated that Arab indexers, when considering the title keywords in the process of indexing Arabic articles, are relying on representative titles.
As far as Arabic topic sentences are concerned, this study revealed that only 60% of the Arab authors include topic sentences in the introduction and objectives parts of their articles. In this technique topic sentences are supposed to be representative of their accompanying articles (see sections 6.3, 6.3.1 and 6.3.2). Since title keywords of Arabic articles were found to be having high matching with keywords found in topic sentences of the same articles, then titles of Arabic articles can be described as having high average of representation of their articles.

9.3.5 Arabic KWOC evaluation

Searchers who used the al-Fihrist type index did not perform better than searchers who used the KWOC type index to find the relevant articles to the nine queries. Although searchers of the al-Fihrist type index achieved better average recall, searchers of the KWOC type index achieved far better precision. Using the T test of significance showed that there is no significant differences between the two categories in terms of retrievability.

It was also observed that searchers who used the KWOC type index to locate the relevant articles were able to achieve 97% of the recall ratio scored by searchers who relied on the al-Fihrist type index. The obtained results also showed that searchers who had used al-Fihrist index before, got an average recall of 73% and an average precision of 86% with 21 minutes as average length time with the al-Fihrist type index compared with an average recall of 71% and an average precision of 98% with 20.8 minutes average time to conduct a search with the KWOC type index. In terms of retrieval effectiveness from these two types of printed indexes to the contents of Arabic periodical articles, the KWOC type is as good as al-Fihrist type. Whatever al-Fihrist is like in terms of effectiveness, this evaluation process
showed that the **KWOC** type index is at least as good as **al-Fihrist** type index in terms of retrieval effectiveness.

It can be said that the use of title keywords of Arabic periodical articles to produce an Arabic **KWOC** type index will lead to the provision of an indexing tool which is as adequate as **al-Fihrist** in terms of retrieval effectiveness, and better in terms of coverage, timeliness and cost effective.

**9.4 Conclusions**

In accordance with the findings from this thesis, the following conclusions may be drawn:

a - The so-called current printed indexes to the contents of Arabic periodical articles are inefficient in terms of coverage, timeliness and cost and deficient in terms of subject headings structure.

b - **Al-Fihrist** index is not any more a current indexing tool and signs of ceasing its publication are looming.

c - The initiative of Dilumun for Publishing in producing two indexing tools with the use of computer does not appear to be as effective as it was intended in being the index of current periodicals which deal with the Islamic matters covered in the Arabic periodicals for the **Islamic Index** and in being the index which covers the Arabic publications on Palestine question and the Arab Israeli conflict for the **Palestine Record**.

d - Titles of academic Arabic periodical articles appear to be informative in 16 disciplines and Arab authors construct informative titles, although they are not faced with the prospect of using the titles of their articles as basis for information retrieval.

e - Titles of academic Arabic periodical articles appear to be representative of their accompanying articles in non-scientific disciplines and
in the field of agriculture.

f - The use of titles of Arabic periodical articles to generate automatic title-derivative printed indexes is not possible currently without human intervention to carry out functions related to the characteristics of the Arabic language. These functions involve the stripping of inseparable prepositions from their following keywords and to control the use of hamazated 'alifs as well as the suppressing of voweling marks.

g - The stripping of inseparable prepositions, the control over the use of hamazated alifs and the suppressing of voweling marks can be carried out during the filling in of input sheets or at the input stage which may be carried out by indexers with typesetting expertise.

h - Since the generation of title-derivative printed indexes by means of computer requires the article titles to be input to the indexing system by the assistance of a human, it is preferable that the same human has the required experience to handle the above-mentioned functions. This policy may be undertaken until software specialists develop packages able to conduct these functions automatically. The function of suppressing voweling marks automatically seems feasible whereas the others seem difficult and require more research.

i - The use of titles of Arabic periodical articles to generate automatic title-derivative printed indexes with human intervention to conduct the functions mentioned above will not contradict the notion of using computers to carry out intellectual and clerical indexing work. Human intervention is required at the input stage, whereas extracting keywords after their comparison against the stored stopwords list and the formatting of the index will be done by computer.

j - Human intervention required in the production of a title keyword
printed index is not a true intellectual intervention as it does not involve title words assessment and keywords extraction from the full article of the processed title.

k - The elimination of voweling marks from titles will not increase the number of homographs or produce vague titles, especially if the index generation depends on article titles which include a limited number of words in comparison with other parts of an article. Also the provision of textual data surrounding the extracted keywords in some title keyword indexing types will play an important role in clarifying the semantic relations among the keywords of a title.

l - Title keyword indexing requires much less indexing effort at the input stage than other indexing systems such as conventional and other similar types.

m - The extra effort required by conventional Arabic indexing does not lead to a significantly improved and better indexing product.

n - This thesis revealed that there are no significant differences at 1% level between the al-Fihrist type index and the KWOC type index in terms of recall and time taken by searchers, but the difference is significant in terms of precision.

o - The generation of title-derivative printed indexes for Arabic periodical articles will suffer from deficiencies arising from synonyms, keyword forms and spelling variations, and the inconsistency in using Arabized terminology. However these problems are not unique to Arabic and procedures taken by Western specialists to alleviate such problems may be considered by Arab specialists in the involved domains.

p - The examination of titles of Arabic periodical articles for the production of Arabic title-derivative printed indexes has been the only sort of
indexing ever investigated for suitability for Arabic periodical articles to overcome deficiencies of the available Arabic printed indexes.

In summary, the findings of this thesis have demonstrated the feasibility and practicability of producing title keyword printed indexes for Arabic articles. Despite some drawbacks which were revealed in this thesis, Arabic title-derivative printed indexes for academic periodical articles appear to be a valid response to the need for earlier dissemination of information relating to the published literature.

9.5 Recommendations

Based upon investigations of the existing techniques for the automatic generation of English title-derivative printed indexes and on the findings of this study, certain practices seem advisable for making Arabic title-derivative printed indexes more practical. However, it is not necessary to repeat what has been recommended for English authors to construct representative titles, because these are mentioned generally throughout the thesis and in several other studies. The following are some recommendations which seem to hold promise for producing better title-derivative printed indexes for the contents of Arabic periodical articles.

a - Editors of Arabic periodicals, associations of Arab writers and other similar concerned groups can influence the adequacy of titles of Arabic articles by setting up instructions to authors for titling their works and by promoting studies which deal with the importance of titles. Instructions may include certain requirements for article titles to be acceptable for publication and by informing authors that the composition of their titles should be based on the possibility of future machine indexing as well as their importance for
information retrieval and the concept of stopwords.

b - Editors of Arabic periodicals are advised to rewrite titles which seem unrepresentative and irrelevant for machine indexing.

d - Authors are advised not to include non-Arabic script and words in their titles, as these will hamper the sorting process and will diminish the space for more representative keywords. Authors who fail to replace non-Arabic words with similar Arabic words should bear in mind that readers may not understand the foreign words.

e - Authors are also advised to refer to countries by their common names and not by the official names which normally occupy more than one word. This procedure will leave more space for the use of relevant keywords.

f - Authors are advised not to use symbols and single letters as these may affect the process of reassembling article titles automatically in the bibliographic sections. Inseparable prepositions which are separated manually at the input stage, may go automatically to their previous shape in the index entry line. To overcome the use of single letters, authors can write them according to their pronunciation. For instance the letter 'alif (א) can be written as (الف) ('alif) to constitute a word.

g - Authors are advised not to use voweling marks. Instead they should try to use terms which do not cause homographs with the absence of voweling marks.

h - It is advised that the Arabic title-derivative indexes adopt the KWOC and its variants types indexing technique which preserve the full title in their index entries, to compensate for probable ambiguity resulting from the elimination of voweling marks. The preservation of full title context increases the accuracy with which index keywords may be interpreted.

i - In this thesis almost all types of title keyword printed indexes used in
the English language were reviewed. However only one type was produced and shown to users. The reason behind the selection of **KWOC** type rather than other types, especially the **KWIC** format, stems from the fact that **KWOC** was proposed in English as an improvement to this type of indexes. It is believed that the use of **KWIC** and all other types which have a similar format and do not provide contextual information may result in some of the title keywords being cut due to the one line space allocation for this type of printed index. One line space allocation for Arabic article titles is not enough, especially if author's names and periodical's names are added to the index entry.

j - The consideration of author's names as keywords would enhance the index and play an assisting role in the information retrieval both on the specific article retrieval level and on the subject retrieval level.

k - As far as the bibliographic information content of the indexed articles is concerned, it is believed that using the **KWOC** type format with the provision of the full bibliographic details would be best, so that Arab users will be required to look up once in the index in order to identify their requested articles.

l - Due to the general drawbacks of the title keyword indexing and to the particular drawbacks of the Arabic title keyword indexing, it is advised to have the title-derivative printed indexes as an additional tool to the contents of current Arabic periodical articles.

9.6 *Future studies*

This research has answered the question of whether titles of Arabic periodical articles can be used as basis for the production of title-derivative printed indexes.
The limitations which surrounded this research were beyond the capacity of the thesis. For instance, one field had to be representative of scientific disciplines (when matching between keywords in titles and their corresponding assigned subject headings), because titles in other fields were inaccessible to the thesis and no indexing tool was available. They may be accessible to researchers based in Arab countries. The variation in the size of samples from one subject to another was imposed, because al-Fihrist index was the only available source of samples and subject headings. Also, the matching process between the occurrence of title keywords in their corresponding topic sentences was done in five subjects due to the inaccessibility of other subjects.

With regards to future studies, this thesis has opened up other questions for which answers are desirable. The use of an Arabic automatic thesaurus to overcome the scattering of related keywords, the automatic generation of cross-references and the automatic spell-checking of keywords are issues for other studies to deal with. Another issue is to what extent can the obtained results be employed for the process of covering Arabic titles in non-Arabic indexes using title keyword technique or producing English title-derivative printed indexes for Arabic articles. Finally, it is left for other studies to investigate the subjects which this thesis did not consider in its examination of titles of Arabic periodical articles. Hopefully when Arabic title indexes have been produced commercially, it will be possible to carry out evaluations of them. This work could also be extended to examine online retrieval of Arabic articles.
Appendices

Appendix 1.1
The Library of Congress transliteration system. This system has been used in slightly modified form in the process of writing the names of Arab authors and the names of the Arabic indexing tools.

<table>
<thead>
<tr>
<th>Arabic Letter</th>
<th>Transliteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>u</td>
<td>u</td>
</tr>
<tr>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td>Long Vowels</td>
<td></td>
</tr>
<tr>
<td>ä</td>
<td>ä</td>
</tr>
<tr>
<td>ü</td>
<td>ü</td>
</tr>
<tr>
<td>Diphthongs</td>
<td></td>
</tr>
<tr>
<td>aw</td>
<td>aw</td>
</tr>
<tr>
<td>ay</td>
<td>ay</td>
</tr>
<tr>
<td>iyy</td>
<td>iyy</td>
</tr>
<tr>
<td>uww</td>
<td>uww</td>
</tr>
</tbody>
</table>

Short Vowels

Long Vowels

Diphthongs
Appendix 1.2

List of the Arabic indexing tools mentioned in the thesis and their English transliteration:

- أبحاث ومقالات مجموعة المجلات العربية التراثية بمكانة الوثائق العربية:
- البibliوغرافيا التحليلية للمقالات الواردة في الدوريات العربية بمكانة:
- البibliوغرافيا الموضوعية العربية: التراثية:
  (Bibliyughrafiya al-Mawdwiyyah al-'Arabiyyah : al-'Arabiyyah)
- البibliوغرافيا الموضوعية العربية: علوم الدين الإسلامي:
  (Bibliyughrafiya al-Mawdwiyyah al-'Arabiyyah : 'Ulm al-Din al-'Islami)
- بحوث مجلات ج. ع. م. العلمية:
  (Buhuth Majallat J. 'A. M. al-'Ilmiyyah)
- بيبليوغرافيا الغرب الإسلامي:
  (Bibliyughrafiya al-Gharb al-'Islami)
- الدليل البibliوغرافي للإنتاج الفكري العربي في مجال المكتبات والتوثيق:
  Dalil al-Bibliyughrafi lil-'Intaj al-Fikri al-'Arabi fi Majal al-Maktabat wal-Tawthiq)
- الدليل البibliوغرافي للإنتاج الفكري العربي في مجال المكتبات والعلومات:
  Dalil al-Bibliyughrafi lil-'Intaj al-Fikri al-'Arabi fi Majal al-Maktabat wal-Ma‘lumat)
- دليل الدوريات الليبية ومحتوياتها:
  (Dalil al-Dawriyyat al-Libiyyah wa-Muhitihatihal-
  Muhtawayatuha)
-ذاكرة الفلسطينية:
  (al-Dhakirah al-Filastiniyyah)
- الصحافة الكويتية في ربع قرن:
  (al-Sahafah al-Kuwaytiyyah fi Rub‘ Qarn)
- كشاف الأهرام:
  (Kashshaf al-'Ahram)
- الكشاف الإسلامي:
  (al-Kashshaf al-'Islami)

326
الكشف التحليلي السنوي لجريدة أم القرى
Jaridat 'Umm al-Qura

الكشف التحليلي للدوريات والنشرات العربية والأجنبية المتواصلة في مركز التوثيق والمعلومات بالإدارة العامة لمجلس التعاون لدول الخليج العربية


Kashshaf al-Tahlili lil-Majallat al- 'Iraqiyyah

الكشف التحليلي للمجلات العراقية

الكشف التحليلي لمقالات الدوريات في الجمهورية العربية المتحدة


Kashshaf al-Darah

Kashshaf al-Dawriyyat al-Sa'udiyyah

Kashshaf al-Dawriyyat al-'Arabiyyah

Kashshaf Risalat al-Maktabah

Kashshaf al-'Arabi

Kashshaf al-Faysal

Kashshaf Majallat Jami'at al-'Imam Muhammad Ibn Sa'ud al-Islamiyyah

Kashshaf al-Majallah al-Tarikhiyyah al-Maghribiya

Kashshaf المقالات في الدوريات السعودية المتخصصة في المكبات والمعلومات

Kashshaf al-Maqalat fi al-Dawriyyat al-Sa'udiyyah al-Mutakhassisah fi al-Maktabat wal-Ma'limat

327
(Kashshaf al-Muqtattaf)

الكشف المقتطف

(Al-Kashshaf al-Mawdu‘i li-
Majallatay Ma‘had al-‘Idarah al-‘Amman)

(Muhtawayat al-Dawriyyat al-‘Arabiyyah)

(Al-Nashrah al-‘Arabiyyah lil-Matbu‘at)

- الكشف الموضوعي لمجلتي معهد الإدارة العامة
- محتويات الدوريات العربية
- النشرة العربية للمطبوعات
Appendix 4.1
List of the Arabic periodicals which were used to study the informativeness of Arabic titles

الآداب
أفاق الاقتصادية
أفاق عربية
الأبحاث
الإدارة العامة
الإداري
الاقتصاد الإسلامي
الاقتصاد العربي
الإنسان والتطور
الباحث
الباحث العربي
تاريخ العرب والعالم
التراث العربي
التراث المستقبلي
التعاون الصناعي في الخليج العربي
تعليم الجماهير
التوثيق الإسلامي
الترشيح
الثورة السكمية
حالات
الحكومية
حفلات الجامعة التونسية
حفلات كلية الآداب (جامعة الكويت)
حولية كلية الإنسانيات والعلوم الاجتماعية
جامعة قطر
حولية كلية الشريعة والدراسات الإسلامية
جامعة قطر
الحياة الثقافية
الدار
دراسات : الاقتصاد والعلوم الإدارية
دراسات : الشريعة والقانون
دراسات : العلوم الإنسانية
دراسات : العلوم العربية
الدراسات الإعلامية للسكان والتنمية والتعمير
دراسات تاريخية
دراسات تربوية (جامعة الملك سعود، كلية
التربيوية)
دراسات عربية
دراسات بعثة
رسالة الخليج العربي
السياسة الدولية
شؤون عربية
شؤون عسكرية
شؤون فلسطينية

الطريق
عالم الفكر
عالم الكتب
الفجر
فصل
فكر
الفكر الإستراتيجي العربي
الفكر العربي
الفكر العربي المعاصر
الفكر العسكري
فنون
قضايا عربية
الكتاب السنوي لعلم الاجتماع
الكتاب العربي
الكرمل
النساء العربي
النساء
المؤرخ العربي
مجلة آثار الاقتصاد الإسلامي
مجلة أبحاث البناء
مجلة الاقتصاد والإدارة
مجلة البحثات الإسلامية
المجلة التاريخية العربية
المجلة التونسية لعلوم التربية
مجلة الجامعة الإسلامية بالمدينة المنورة
مجلة جامعة الإمارات العربية المتحدة
المجلة الجغرافية السعودية
المجلة العربية العالمية
مجلة الحقوق
مجلة دراسات الخليج والجزيرة العربية
المجلة الطبية الأردنية
المجلة الطبية السعودية
المجلة العربية للإدارة
المجلة العربية للبحوث التاريخية
المجلة العربية للتراث
المجلة العربية للدراسات اللغوية
المجلة العربية للدفاع الاجتماعي
المجلة العربية للعلوم
المجلة العربية للعلوم الإنسانية
المجلة العربية للعلوم الاجتماعية
مجلة العلوم الاجتماعية

329
مجلة العلوم القانونية والسياسية
مجلة العلوم الهندسية (جامعة الملك سعود)
مجلة العلم العربية
المجلة القانونية التونسية
مجلة كلية الآداب (جامعة الملك سعود)
مجلة كلية الآداب والعلوم الإنسانية في جامعة الملك سعود
مجلة كلية الزراعة (جامعة الملك سعود)
مجلة كلية الشريعة وأصول الدين بالقصيم (جامعة الإمام محمد بن سعود)
مجلة كلية الشريعة والدراسات الإسلامية
مجلة كلية الشريعة والدراسات الإسلامية بالاحساء
مجلة كلية الشريعة واللغة العربية، أبيها
مجلة كلية العلوم (جامعة الملك سعود)
مجلة كلية العلوم الاجتماعية
مجلة كلية الفقه
مجلة المجمع العلمي العراقي
مجلة مجمع اللغة العربية الأردن
مجلة مجمع اللغة العربية بدمشق
مجلة مركز البحوث (جامعة محمد بن سعود)
مجلة المسلم المعاصر
المجلة المصرفية للقانون الدولي
مجلة معهد البحوث والدراسات الإسلامية
مجلة معهد البحوث والدراسات العربية
المجلة المغربية للتوثيق
المجلة العربية للقانون واقتصاد التنمية
المستقبل العربي
المعرفة
المدارة
المналح
المتتبع
مواقف
الناشر العربي
النفط والتنمية
النقل البري العربي
الموقع
الموحدة
### Appendix 4.2

The stopwords list which was used to study the information content of titles of Arabic periodical articles

<table>
<thead>
<tr>
<th>English</th>
<th>Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>connections</td>
<td>ارتباطات</td>
</tr>
<tr>
<td>instructions</td>
<td>إرشادات</td>
</tr>
<tr>
<td>using</td>
<td>استخدام</td>
</tr>
<tr>
<td>usage</td>
<td>إستعمال</td>
</tr>
<tr>
<td>reexamination</td>
<td>إعادة اكتساب</td>
</tr>
<tr>
<td>reconsideration</td>
<td>إعادة اعتبار</td>
</tr>
<tr>
<td>reconsiderations</td>
<td>إعادة اعتبارات</td>
</tr>
<tr>
<td>reinvestigation</td>
<td>إعادة التحقق</td>
</tr>
<tr>
<td>reinvestigations</td>
<td>إعادة التحققات</td>
</tr>
<tr>
<td>reassessment</td>
<td>إعادة تحسين</td>
</tr>
<tr>
<td>reappraisal</td>
<td>إعادة تقدير</td>
</tr>
<tr>
<td>considerations</td>
<td>اعتبارات</td>
</tr>
<tr>
<td>utilizing</td>
<td>الإفادة من</td>
</tr>
<tr>
<td>suggestion</td>
<td>إقتراح</td>
</tr>
<tr>
<td>suggestions</td>
<td>إقتراحات</td>
</tr>
<tr>
<td>to</td>
<td>إلى</td>
</tr>
<tr>
<td>examination</td>
<td>إختبار</td>
</tr>
<tr>
<td>examinations</td>
<td>إختبارات</td>
</tr>
<tr>
<td>reflections</td>
<td>إبعاسات</td>
</tr>
<tr>
<td>finding</td>
<td>إيجاد</td>
</tr>
<tr>
<td>in</td>
<td>ب</td>
</tr>
<tr>
<td>toward, towards</td>
<td>باتجاه/ نحو</td>
</tr>
<tr>
<td>particularly</td>
<td>بالذات</td>
</tr>
<tr>
<td>with respect to</td>
<td>بال نسبة</td>
</tr>
<tr>
<td>research</td>
<td>بحث</td>
</tr>
<tr>
<td>without</td>
<td>بدون</td>
</tr>
<tr>
<td>after</td>
<td>بعد، إثر</td>
</tr>
<tr>
<td>some</td>
<td>بعض</td>
</tr>
<tr>
<td>including</td>
<td>بماض، ما فيه، ب ما</td>
</tr>
<tr>
<td>against</td>
<td>بوجهة، ضد</td>
</tr>
<tr>
<td>by</td>
<td>بواسطة، عبر</td>
</tr>
<tr>
<td>among, between</td>
<td>بين</td>
</tr>
<tr>
<td>experiments</td>
<td>تجارب</td>
</tr>
<tr>
<td>under</td>
<td>تحت</td>
</tr>
<tr>
<td>investigations</td>
<td>تحريات</td>
</tr>
<tr>
<td>inquiry</td>
<td>استفسال</td>
</tr>
<tr>
<td>inquiries</td>
<td>استفسالات</td>
</tr>
<tr>
<td>applications</td>
<td>تطبيق</td>
</tr>
<tr>
<td>applied</td>
<td>تطبيقي</td>
</tr>
<tr>
<td>other</td>
<td>آخر</td>
</tr>
<tr>
<td>opinions</td>
<td>آراء</td>
</tr>
<tr>
<td>aspects</td>
<td>أفاق</td>
</tr>
<tr>
<td>researches</td>
<td>بحوث</td>
</tr>
<tr>
<td>dimensions</td>
<td>أبعاد</td>
</tr>
<tr>
<td>further</td>
<td>أبعد، أكثر</td>
</tr>
<tr>
<td>parts</td>
<td>أجزاء</td>
</tr>
<tr>
<td>better</td>
<td>أفضل</td>
</tr>
<tr>
<td>situation</td>
<td>حالة</td>
</tr>
<tr>
<td>an</td>
<td>اذاء تفكير</td>
</tr>
<tr>
<td>a</td>
<td>اذاء تفكير</td>
</tr>
<tr>
<td>means</td>
<td>أدوات</td>
</tr>
<tr>
<td>questions</td>
<td>استفسال</td>
</tr>
<tr>
<td>techniques</td>
<td>أساليب</td>
</tr>
<tr>
<td>reasons</td>
<td>أسباب</td>
</tr>
<tr>
<td>technique</td>
<td>أساليب</td>
</tr>
<tr>
<td>shapes</td>
<td>أشكال</td>
</tr>
<tr>
<td>works</td>
<td>أعمال</td>
</tr>
<tr>
<td>things</td>
<td>أغراض</td>
</tr>
<tr>
<td>better</td>
<td>أفضل</td>
</tr>
<tr>
<td>more</td>
<td>أكثر</td>
</tr>
<tr>
<td>the</td>
<td>ال</td>
</tr>
<tr>
<td>or</td>
<td>أو</td>
</tr>
<tr>
<td>before</td>
<td>أمام</td>
</tr>
<tr>
<td>examples</td>
<td>أمثلة</td>
</tr>
<tr>
<td>themselves</td>
<td>أنفسهم</td>
</tr>
<tr>
<td>features</td>
<td>أنماط</td>
</tr>
<tr>
<td>types</td>
<td>أنواع</td>
</tr>
<tr>
<td>varieties</td>
<td>تنوعات</td>
</tr>
<tr>
<td>or</td>
<td>أو</td>
</tr>
<tr>
<td>descriptions</td>
<td>وصف</td>
</tr>
<tr>
<td>situation</td>
<td>وضع</td>
</tr>
<tr>
<td>or, which</td>
<td>أي</td>
</tr>
<tr>
<td>where</td>
<td>ما، حيث</td>
</tr>
<tr>
<td>destinations</td>
<td>إنجهارات</td>
</tr>
<tr>
<td>two</td>
<td>إثنان</td>
</tr>
<tr>
<td>replies</td>
<td>إجابات</td>
</tr>
<tr>
<td>procedure</td>
<td>إجراء</td>
</tr>
<tr>
<td>procedures</td>
<td>إجراءات</td>
</tr>
<tr>
<td>experiment</td>
<td>اختبار</td>
</tr>
<tr>
<td>experiments</td>
<td>إختبارات</td>
</tr>
<tr>
<td>connection</td>
<td>ارتباط</td>
</tr>
</tbody>
</table>

331
supports
role
dek
that, an
view
R"j
relationship
ربما
question
سؤال
would
would
policies
applied
across, via, through
contrary, opposite
relation
علاقة
on, over, upon
على, عليها, على
thereon
عليه، بعد ذلك
from
عن
elements
عناصر
at
غير
up
فوق
in, at
في
following
في أعقاب
concerning
فيما يخص
applicable
قابل للتطبيق
before
قبل
near
قرب
issues
قضايا
issue
قضية
was
كان
were
كانت
very, much
كثير
every
كل
as, like
كما
how
كيف
for
لأجل
no
للاخص
for some
لبعض
for
لدى
why
لم
when

dealings
definition
commentary
comments
commendations
commendations
details
illustration
illustrations
reports
test report
evaluation
exercises
exercise
preliminary
empirical
employing
to implement, to use
which, that
again
ثاني
part
جزء
all
جميع
aspects
characteristics
aspect
جهة
aspects
aspects
cases
cases
cases
study
until, even, till
the acquisition on
solutions
solutions
circa
حوالى (للتاريخ)
about, around
about
out
عولج
special
خصائص
especially
خصائص
our
خاصيتنا
his
خاصيته
its
خاصية (الأعمال)
his
خاصيتها
their
خاصيتهم
experiences
خبرات
experience, expertise
surveys, summaries
summaries
survey, summary
survey, summary
during, within
during, within
within, inside
within, inside
studies
studies
studies
Appendix 4.3
List of the English periodicals which were used to study the informativeness of English titles

The Agricultural History Review
American Anthropologist
The American Historical Review
American Journal of Sociology
Annals of the Association of American Geographers
Aslib Proceedings
Australian Geographical Studies
Australian Geographical Studies
Australian Journal of Agricultural Research
British Journal of Educational Studies
Bulletin of Economic Research
The British Journal of Social Work
The Cambridge Law Journal
The English Historical Review
European Economic Review
European History Quarterly
French Historical Studies
General Linguistics
Harvard International Law Journal
Higher Education
The Historical Journal
The International History Review
International Journal of Comparative Sociology
International Library Review
International Philosophical Quarterly
Journal of Development Economics
The Journal of Economics
Journal of Humanistic Psychology
Journal of Information Science
Journal of International Economics
The Journal of Modern History
Journal of Organic Chemistry
Journal of Physics
Lancet
The Law Quarterly Review
Legal Studies
Linguistics
Personality and Social Psychology Bulletin
Personnel Psychology
The Philosophical Quarterly
Philosophy
The Political Quarterly
Political Science Quarterly
Proceedings of the Institute of Mechanical Engineers
The Review of Economic Studies
Studies in Higher Education

334
Appendix 5.1
List of the Arabic periodicals used in the matching process between titles of Arabic articles and their corresponding subject headings

المؤرخة العربية
المجلة الثقافية
المجلة العربية لبحوث التربية
المجلة العربية للبحث التربوي
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس والدراسات الإسلامية
المجلة العربية للدروس و
Appendix 6.1
List of the Arabic phrases found at the beginning of the examined Arabic topic sentences.

- البَحْث الذي بين ابْدِنَا يَلْقِي الضَّرْعٌ على
- The research in hands throws the light on the matter of...

- هذه دراسة عن
- This is a study on...

- في هذه المقالة سنستثمن البَحْث موضوع
- In this article we will have a general view on...

- In this article I will attempt to...
- In this article I aim to...

- يَهْدِف هذا البحث إلى
- This research aims to...

- يَتَبَنَّى هذا المقال قضية
- This article discusses the matter of...

- تحاول هذه الدراسة تحضيل
- This study attempts to analyse the...
Appendix 6.2
List of the Arabic periodicals from which the samples of titles and topic sentences were selected

البحوث
الإدارة العامة
الاقتصاد
الباحث
البيان
المحاضر
حوليات الجامعة التونسية (كلية الآداب)
حوليات المكتبات والمعلومات
دراسات العلوم الإنسانية
دراسات العلوم الزراعية
دراسات الخليج والجزيرة العربية
رسالة الكتب
شؤون عربية
عالم الفكر
عالم الكتب
الفكر العربي
السادات
المجلة البحوث والدراسات العربية
المجلة دراسات الخليج والجزيرة العربية
المجلة العربية للعلوم الإنسانية
المجلة العربية للمعلومات
المجلة العلمية الاجتماعية
مجلة كلية الآداب (جامعة الرباط)
مجلة كلية الآداب (جامعة الملك سعود)
المستقبل العربي
العربية
مكتبة الإدارة
مواقف
الموقف العربي
النفط والتعاون العربي
النفط والتنمية
الوحدة
النظام

337
Appendix 7.1
Copy of the form which was distributed to the searchers of al-Fihrist type index.
(the original form did not include English translation)

This form is intended to obtain information on the performance of Arabic periodical indexes. The answers will be analysed in a chapter of a doctoral thesis about the indexing of Arabic periodical articles.

(Dear colleague,
Thank you very much for your help, and I wish you the best of luck with your study and work.
Before I invite you to answer the queries, I would like to explain to you the nature of the attached index. It is a general subject index for a number of Arabic periodical articles and known as the subject index of periodical articles. It consists of the subject headings assigned by the indexer, who selected them from controlled Arabic subject heading lists, to express the contents of the articles included in the index.
The main entries in this index are the subject headings, which are arranged in accordance with the Arabic alphabet. When answering, would you please cite the article title or the index entry number under each query).
(user number) 

(搜索时间) 

(ایپا الیمعل : ارجو ان تجب علی السؤالین التالیین) 

(دمه البحث) 

(سمع) (نه) 

(وئف) (نعم)

(هل استعملت كشاف (فهرسا) للدوريات باللغة NON-العربية من قبل؟) 

(هل استعملت كشاف (فهرسا) للدوريات باللغة العربية من قبل؟) 

(نام ما هو اسم ذلك الكشاف (أو تلك الكشافات) ) 

(في حال كان الجواب نعم ما هو اسم ذلك الكشاف (أو تلك الكشافات) )

(بسم الله الرحمن الرحيم)

(لا) (نعم)

(لا) (نعم)

(لا) (نعم)

(لا) (نعم)
1. Look for articles on the Islamic concept of economy

2. Look for articles on combating illiteracy in Bahrain

3. Find articles on teaching in the Gulf region
(look for articles on the renaissance period)

(look for articles on banks)

(look for articles about children)
- إبحث عن المقالات المتعلقة باليهودية
(find articles about Judaism)

- إبحث عن مقالات حول ابن رشد
(find articles about Ibn Rushd)

- إبحث عن مقالات حول الدولة السورية
(find articles about the State of Syria)
Appendix 7.2
Copy of the form which was distributed to the searchers of the Arabic KWOC type index.
(the original form did not include English translation)

This form is intended to obtain information on the performance of Arabic periodical indexes. The answers will be analysed in a chapter of a doctoral thesis about the indexing of Arabic periodical articles.

Dear colleague,

Thank you very much for your help, and I wish you the best of luck with your study and work.

Before I invite you to answer the queries, I would like to explain to you the nature of the attached index. It is known as title words index and consists of the keywords found in the titles of some Arabic periodical articles.

In contrast with stopwords such as "in, about, to" and the like, the keywords are the ones which, in themselves, convey certain meanings.

The main entries in this index are the keywords, which are arranged in accordance with the Arabic alphabet. When answering, would you please cite the article title or the index entry number under each query).
Dear colleague: I would like you to answer the following two questions.

1. Have you used a non-Arabic periodical index before?
   - (No) ☐ (Yes) ☐
   
   If the answer is yes, what is the name of that index or those indexes?

2. Have you used an Arabic periodical index before?
   - (No) ☐ (Yes) ☐
   
   If the answer is yes, what is the name of that index or those indexes?
1 - إبحث عن المقالات التي تتحدث عن الاقتصاد الإسلامي
(look for articles on the Islamic concept of economy)

2 - إبحث عن المقالات التي تتحدث عن محو الأمية في البحرين
(look for articles on combating illiteracy in Bahrain)

3 - إبحث عن المقالات التي تتحدث عن التعليم في دول الخليج
(find articles on teaching in the Gulf region)
4 - Look for articles on the renaissance period

5 - Look for articles on banks

6 - Look for articles about children
7 - Find articles about Judaism

8 - Find articles about Ibn Rushd

9 - Find articles about the State of Syria
<table>
<thead>
<tr>
<th>Name</th>
<th>Uniform value</th>
<th>Write a full vowel letter</th>
<th>Write a full digraph letter</th>
<th>With both</th>
</tr>
</thead>
<tbody>
<tr>
<td>آن</td>
<td>۰</td>
<td>۰</td>
<td>۰</td>
<td>۰</td>
</tr>
<tr>
<td>د</td>
<td>۱</td>
<td>۱</td>
<td>۱</td>
<td>۱</td>
</tr>
<tr>
<td>ت</td>
<td>۲</td>
<td>۲</td>
<td>۲</td>
<td>۲</td>
</tr>
<tr>
<td>۳</td>
<td>۳</td>
<td>۳</td>
<td>۳</td>
<td>۳</td>
</tr>
<tr>
<td>ح</td>
<td>۴</td>
<td>۴</td>
<td>۴</td>
<td>۴</td>
</tr>
<tr>
<td>۵</td>
<td>۵</td>
<td>۵</td>
<td>۵</td>
<td>۵</td>
</tr>
<tr>
<td>۶</td>
<td>۶</td>
<td>۶</td>
<td>۶</td>
<td>۶</td>
</tr>
<tr>
<td>را</td>
<td>۷</td>
<td>۷</td>
<td>۷</td>
<td>۷</td>
</tr>
<tr>
<td>سر</td>
<td>۸</td>
<td>۸</td>
<td>۸</td>
<td>۸</td>
</tr>
<tr>
<td>۹</td>
<td>۹</td>
<td>۹</td>
<td>۹</td>
<td>۹</td>
</tr>
<tr>
<td>۱۰</td>
<td>۱۰</td>
<td>۱۰</td>
<td>۱۰</td>
<td>۱۰</td>
</tr>
<tr>
<td>۱۱</td>
<td>۱۱</td>
<td>۱۱</td>
<td>۱۱</td>
<td>۱۱</td>
</tr>
<tr>
<td>۱۲</td>
<td>۱۲</td>
<td>۱۲</td>
<td>۱۲</td>
<td>۱۲</td>
</tr>
<tr>
<td>۱۳</td>
<td>۱۳</td>
<td>۱۳</td>
<td>۱۳</td>
<td>۱۳</td>
</tr>
<tr>
<td>۱۴</td>
<td>۱۴</td>
<td>۱۴</td>
<td>۱۴</td>
<td>۱۴</td>
</tr>
<tr>
<td>۱۵</td>
<td>۱۵</td>
<td>۱۵</td>
<td>۱۵</td>
<td>۱۵</td>
</tr>
<tr>
<td>۱۶</td>
<td>۱۶</td>
<td>۱۶</td>
<td>۱۶</td>
<td>۱۶</td>
</tr>
<tr>
<td>۱۷</td>
<td>۱۷</td>
<td>۱۷</td>
<td>۱۷</td>
<td>۱۷</td>
</tr>
<tr>
<td>۱۸</td>
<td>۱۸</td>
<td>۱۸</td>
<td>۱۸</td>
<td>۱۸</td>
</tr>
<tr>
<td>۱۹</td>
<td>۱۹</td>
<td>۱۹</td>
<td>۱۹</td>
<td>۱۹</td>
</tr>
<tr>
<td>۲۰</td>
<td>۲۰</td>
<td>۲۰</td>
<td>۲۰</td>
<td>۲۰</td>
</tr>
<tr>
<td>۲۱</td>
<td>۲۱</td>
<td>۲۱</td>
<td>۲۱</td>
<td>۲۱</td>
</tr>
<tr>
<td>۲۲</td>
<td>۲۲</td>
<td>۲۲</td>
<td>۲۲</td>
<td>۲۲</td>
</tr>
<tr>
<td>۲۳</td>
<td>۲۳</td>
<td>۲۳</td>
<td>۲۳</td>
<td>۲۳</td>
</tr>
<tr>
<td>۲۴</td>
<td>۲۴</td>
<td>۲۴</td>
<td>۲۴</td>
<td>۲۴</td>
</tr>
<tr>
<td>۲۵</td>
<td>۲۵</td>
<td>۲۵</td>
<td>۲۵</td>
<td>۲۵</td>
</tr>
<tr>
<td>۲۶</td>
<td>۲۶</td>
<td>۲۶</td>
<td>۲۶</td>
<td>۲۶</td>
</tr>
<tr>
<td>۲۷</td>
<td>۲۷</td>
<td>۲۷</td>
<td>۲۷</td>
<td>۲۷</td>
</tr>
<tr>
<td>۲۸</td>
<td>۲۸</td>
<td>۲۸</td>
<td>۲۸</td>
<td>۲۸</td>
</tr>
<tr>
<td>۲۹</td>
<td>۲۹</td>
<td>۲۹</td>
<td>۲۹</td>
<td>۲۹</td>
</tr>
<tr>
<td>۳۰</td>
<td>۳۰</td>
<td>۳۰</td>
<td>۳۰</td>
<td>۳۰</td>
</tr>
</tbody>
</table>

Appendix 8.1: The Arabic characters and their various shapes within a word. Taken from Wright (1933)

348
Reference List


Abdul-Huq, A.M. see ‘Abd al-Haqq, A. M.


Bachir, Imad A.M., see Bashir, ‘Imad A.M.


Al-Bustani, 'Idwar A., 1982. Courses in informational composition: a chapter in writing journalistic titles (دروس في الإنشاء الإعلامي , فصل في كتابة العناوين الصحفي). Beirut: the Lebanese University, Faculty of Information and Documentation. (in Arabic)


Collison, Robert L., 1959. Indexes and indexing: guide to the indexing of books and collection of books, periodicals, music gramophone records, films and other material, with a reference section and suggestions for further reading. 2nd ed. London: Ernest Benn Ltd.

Contents of Arabic Periodicals (Cyprus), 1989-


Dobis/Libis at Liverpool University Library, 1986. Vine, 64, pp. 3-14.

Farsuni, Fu’ad, 1981. The professional structure of the Arabic subject heading list. preparation by the Cataloguing and Classification Department at Riyadh University. "البنية الفنية لقائمة رؤوس الموضوعات العربية. إعداد قسم الفهرسة والتصنيف بجامعة الرياض." World of Books (عالم الكتب), 2 (1), pp. 32-42. (in Arabic)


Frost, Carolyn O., 1989. Title words as entry vocabulary to LCSH: correlation between assigned LCSH terms and derived terms from titles in bibliographic records with implications for subject access in an online catalogs. Cataloging and Classification Quarterly, 10 (1,2), pp. 165-179.


Al-Hajrasi, Sa‘d Muhammad, 1974b. Al-‘Ahram Index amongst the bibliographic periodicals in the Arab World (الбиблиوغرافيا بالوطن العربي). In: Al-‘Ahram Index 1974. Cairo: Centre for Organization and Microfilm in al-‘Ahram (مركز التنظيم والبيكروفيلم في الأهرام), pp. 1-47. (in Arabic)


'Itayyim, Mahmud 'Ahmad, 1986. The problems of indexing and abstracting in Arabic language (مشكلات التكشيف والاستخلاص في اللغة العربية). In : the 9th Training Session for Arab Documentalists about Indexing and Abstracting : procedures and services. Tunisia : ALDOC, pp. 123-183 (in Arabic)


Kasem, Hishmat M.A., see Qasim, Hishmat M. A.


Musa, 'Abd al-Mun'im Muhammad, 1973. Preparation of indexes: concise memorandum (إعداد الكشافات: مذكرة موجزة) - Cairo: Cairo University, Computer Centre.(in Arabic)


Riyadh University; Deanship of Libraries Affairs; Cataloguing and Classification Department, 1978. Arabic Subject Headings (رئوس الموضوعات العربية). Supervised by Nasir al-Suwaydan. Riyadh: Deanship of Libraries Affairs. (in Arabic)


Rzuwqi, Na‘imah H., 1985. The use of computers in formatting and producing periodical indexes (إستخدام الحاسبات الإلكترونية في إخراج وإنتاج كشافات الدوريات). (Documentation Series, no. 5). Baghdad : Gulf States Information Documentation Centre. (in Arabic)


Al-Suwaydi, 'Ali al-Sulayman, 1988. Permuted indexes and information retrieval in the Arabic language (الاكتشافات التبادل وإسترجاع المعلومات في اللغة العربية ( Series no. 2). Riyadh : King Fahd Library Publications (in Arabic)


Ubaydli, Ubaidly, see 'Ubaydli, 'Ubaydli,


Wright, W., 1933. A grammar of the Arabic language. Cambridge: Cambridge University Presss.


Yusuf, Qasim, 1989. Reversed subject headings are better than the natural subject headings (رؤوس الموضوعات العربية المقلوبة أفضل من الرؤوس الطبيعية). World of Books (عالم الكتب), 10 (4), 198-503. (in Arabic)

