

Distributed Search in National electronic Library for Health

Patty Kostkova
Institute of Health Sciences
The City University
London, UK
patty@soi.city.ac.uk

Jane Mani-Saada
Institute of Health Sciences
The City University
London, UK
J.Mani-Saada@city.ac.uk

Julius R Weinberg
Institute of Health Sciences
The City University
London, UK
J.R.Weinberg@city.ac.uk

Abstract

Modern healthcare specialists are overwhelmed by the volume of medical information available on the Internet. However, they cannot always find what they need when they need it. The UK National electronic Library for Health (NeLH) is addressing this issue by providing a single-entry portal to evidence-based medical information on the Internet enhanced with a quality tag assigned by professional experts.

In order to fully utilize the potential of an Internet-based library, the NeLH is distributed and consists of a number of Virtual Branch Libraries (VBLs), each dedicated to a particular disease or a medical area. VBLs are dynamically updated and their design reflects the needs of each particular user base. However, users accessing a single VBL may want to search the entire NeLH or should have the option of being able to search the entire NeLH. Therefore, support for a distributed search according to an adopted topology of VBL servers is essential. In this paper, we present an agent-based solution to distributed search based on a star topology of the NeLH.

1. Introduction

The National electronic Library for Health Programme (NeLH) [1] is developing a digital library for NHS staff, patients and the public in the UK. This will provide a gateway to the evidence-based medical knowledge on the Internet.

Modern healthcare specialists as well as general public are overwhelmed with medical information available on the Internet. However, they cannot find a piece of information when and where they need it [1]. Therefore, the Department of Health (DoH) UK have proposed the development of a National electronic Library for Health (NeLH) [2].

One of the crucial requirements of the NeLH is providing the *best available evidence* and *quality-tagging* the medical information for all user groups. This will distinguish the NeLH from many other existing medical sites (such as Medline [3]) which typically offer comprehensive searches of bibliographic databases without giving the user an insight into the level of evidence and quality of the presented information [4]. Furthermore, the NeLH is dedicated to provide quality tagged information related directly to patient care and does not address directly the science literature.

This paper focuses on technical aspects of the distribution of the library into the Virtual Branch Libraries (VBLs), each dedicated to a specific area or a disease. There could conceivably be areas of overlap for example between the VBL for communicable disease and the VBLs for public health or primary care. The need for distributed search across the NeLH is essential in order to allow users to fully utilize the Internet-based resource. However, there is a tradeoff between flexibility and implementation simplicity.

On one hand, the independence of each VBL with control over technical aspects and user interface design could ensure that the specific needs of each user group are supported. On the other hand, a distributed solution brings an implementation complexity, communication overhead and is not viable for financial reasons.

In this paper, we discuss how Intelligent Agents are implemented in searching information in the Internet [5], and focus on the key issue of this paper – the distributed search and topology of the distributed VBL libraries.

This paper starts by giving a brief overview of the NeLH project (section 2), and then focuses on the structure of the NeLH (section 3). A distributed search is discussed in section 4, followed by a discussion of the

adopted approach, in section 5. Finally, in section 6 we discuss the project status, and in section 8 we conclude.

2. Project Background

The three major goals of the NeLH are [6]:

- To provide health care professionals and the public with knowledge and know-how to support health care related decisions
- To provide easy access to best current knowledge and know-how on the Internet
- To improve health and health care, clinical practice and patient choice.

The main source of medical evidence are books, journals, and Internet-based sources. These include: Public Health Laboratory Service [7], Cochrane database [8], NHS Centre for Reviews and Dissemination, Effective Health Care Bulletins, British National Formulary, Centers for Disease Control, British Medical Journal, and others.

However, the quality, reliability and "non-biasness" of provided information significantly vary. For example, studies of Mulrow [9], Oxman and Guyatt [10] have revealed how unreliable some editorials and review articles can be if they are not prepared systematically. In particular, although readers rely on journal review articles and editorials, the scientific evidence of these is inherently unreliable and biased towards a positive and optimistic view of the effectiveness of intervention [11].

Therefore, in order to overcome the problem of bias and unreliable medical evidence, and to ensure that the best current knowledge is delivered, NeLH aims to fully support evidence-based healthcare [12]. However, in order to achieve this goal the available information must be kept up-to-date, the documents review process needs to be managed and automated and users specific needs should be reflected.

3. Structure of the NeLH

In this section, we will look at the general issues of the NeLCD digital library and discuss internal data representation.

3.1. Overview of the library

The NeLCD Knowledge Base is now a dominant commodity and knowledge distribution is so vital that the contribution made by computing, telecommunications and the World Wide Web are obviously of increasing importance. The development of knowledge management within healthcare is both necessary and possible [13]. It is based on existing medical evidence-based sites, such as PubMed, Cochrane Database, etc.

The core content of the NeLH can be likened to a central reference library and Virtual Branch Libraries to local branches, which serve the needs of the communities in which they based. So far, a number of prototype VBLs are being developed, such as Cancer, Child Health, Communicable Diseases, Diabetes, Diagnosis, Emergency Care, Health Informatics, Heart Disease, and others.

Our team is responsible for development of the NeLCD [14], the Communicable Disease branch of the NeLH, which addresses issues in prevention, investigation treatment and control of communicable disease. This is a single information gateway, a portal, to evidence-based information related to communicable disease with respect to all user groups – clinicians, GPs, public health professionals, environmental health officers, infectious control nurses, general public and others.

Figure 1 illustrates the NeLH framework. NeLCD stands for the National electronic Library for Communicable Disease (as discussed above), NeLC is the National electronic Library for Cancers, and the NeLPC is the National electronic Library for Primary Care.

NeLH Structure

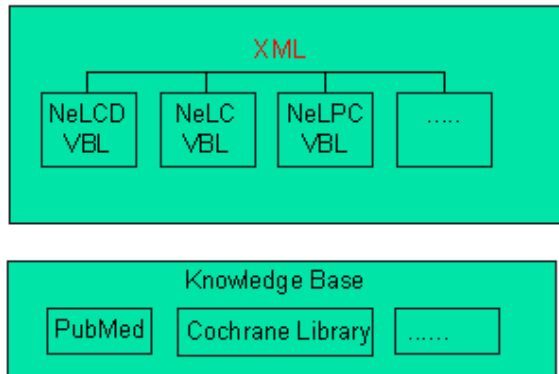


Figure 1: The NeLH Structure

3.2. Data Representation

The NeLH is to store evidence-based medical information on clinical and non-clinical subjects. However, there is no common agreement on ontology, or agreed standards in health care (coding standards, data representation standards and common legal and ethical recommendations). For example, there is no common internationally accepted clinical coding scheme – currently, several coding systems are being used by different organizations and for different purpose (clinical coding, librarianship, etc.) These include, MESH, CTLV3/SNOMED and ICD10. This is not only a UK but also an international issue.

In order to support a user-customizable search, documents in the library need to be precisely described. The Dublin Core Metadata initiative (<http://www.purl.org/DC>) defines a list of fields characterizing an electronic document for cataloging and search purposes. The NeLH adopted and extended this framework to better meet the requirements of quality and "up-to-date-ness". Issues related to metadata are covered in greater detail in the paper by Kamel Boulos describing a preparatory study of metadata, RDF and problem-knowledge coupling for the NeLH. [15]. Every document in NeLH will be described by the following fields - called "electronic catalogue card".

Dublin Core Field	Description
<i>Title</i>	<i>Title of the document</i>
<i>Creator</i>	<i>Author(s) of the document</i>
<i>Subject</i>	<i>Keywords for indexing</i>
<i>Publisher</i>	<i>The publishing organization or the Internet site</i>
<i>Date - publishing - posting - expiry</i>	<i>Date of publication where stated or date accessed if it is an internet site Date of NeLCD posting Date for review</i>
<i>Type</i>	<i>Publication type</i>
<i>Format</i>	<i>Software format</i>
<i>Identifier</i>	<i>Identifier (ISBN, URL)</i>
<i>Source</i>	<i>Document Bibliography</i>
<i>Language</i>	<i>Document Language</i>
<i>Relation</i>	<i>Reference to related documents</i>
<i>Coverage</i>	<i>Medical "category"</i>
<i>Quality Tag</i>	<i>Level of evidence</i>
<i>Check List</i>	<i>Answers to Check List Questions</i>
<i>Description</i>	<i>"bottom line" summarizing the document</i>
<i>ID</i>	<i>Unique identifier</i>

Table 1: The Electronic Catalogue Card

3.3. The Appraisal Process

As each VBL serves specific group of healthcare professionals with particular information needs, there is a need for variations in the document appraisal process to meet the specific needs of each VBL's user base.

In this section, we will describe the appraisal process as it is set up within the NeLCD. The model of collaboration with professionals involved in the appraisal process may vary among VBLs, however, the key issues regarding the support for evidence and quality-tagging are agreed by all VBL teams.

As only pre-processing and quality-tagging of available information before incorporating it into NeLCD could ensure the required quality of the site, the core of NeLCD will consist only of appraised documents which have passed the NeLCD editorial procedure. This does not restrict the physical location of the actual documents – they could be local or accessed at their original source. Nor does it restrict the type of document available – the NeLCD attempts to present the “best available evidence”, this could be a meta-analysis, or where there is little literature, a case report (but this is clearly indicated by the Level of Evidence, that is the Quality tag in the Dublic Core definition).

As all information posted to NeLCD must be kept up-to-date, it is essential to review the core material regularly, even when no contradicting evidence has been found. For this reason every document has to go through a review process when its “Expiry date” (see Table 1) has passed. This is usually one year from the date of posting and then annually thereafter. Each document considered the best available evidence on a particular subject is assigned a *quality tag*, by a member of a professional society or expert group.

3.4. Agent-based search in the NeLCD

There are currently four basic agent concepts in the NeLCD. First two, Intelligent Search Agents and Pro-active Alert Agents, are involved in the search process and user profiling and customization. The later, Reactive Review Agents and Reactive Expiry Agents, are in charge of various aspects of the library review process. The NeLCD is illustrated in Figure 2.

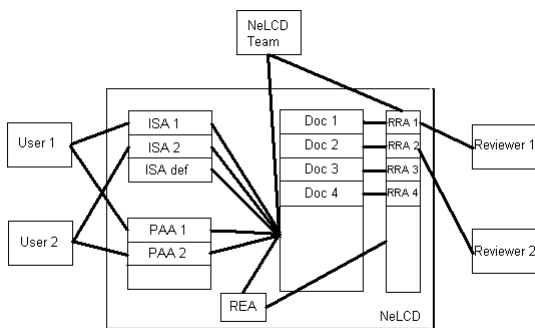


Figure 2: Agents in NeLCD

The primary goal of the library is to provide a single portal for searching for the best, up-to-date, evidence-based medical information available. Special Intelligent User Interface (IUI) agents [16], Intelligent Search Agents (ISA), are used to provide users with a user-friendly adaptive interface to define a query for searching medical information in the library.

In addition to defining keywords for a query and their combination by logical operators, users can customize the search by restricting the resultant set to certain criteria, as indicated by the Dublic Core fields. For example, the agent can search for all documents regarding meningitis and children and display only Randomized Control Trials, or Case Studies.

Pro-active Alert agents (PAA) are performing tasks on users-behalf, such as alerts according to user interest.

Reactive Review Agents (RRA) and the Reactive Expiry Agent (REA) are responsible for autonomous management of the review process.

The details of ensuring that the site is kept up-to-date using Intelligent Agents is discussed in detail by Kostkova in the paper on agent-based data management [17].

4. Distributed Searching in NeLH

The above section discussed technical aspects of the search for medical information within the NeLCD. Each VBL team is responsible for implementing the search facility within their virtual branch library, while accomplishing the requirements on data representation using the agreed Dublin Core-based electronic catalogue card. These are exchanged in XML format.

Medical ontology of data exchange and technical aspects of the interoperability among VBLs and the actual topology of VBLs are investigated in this section.

4.1. NeLH Topology

The NeLH architecture is being designed not as a simple client-server, but a set of VBL servers communicating with users who wants to search all available information within the NeLH. Therefore, the first technical issue to enable a distributed search is the topology of the digital library.

We are investigating a *star topology*, that is, a VBL processing a search will contact the NeLH server, which acts as a mediator, to obtain search results from all other VBLs. Another solution would be to implement a *Peer-to-Peer* communication where each VBL would separately contact every other VBL to receive all potential results. We opted for the star solution for its implementation simplicity and easier consistency maintenance.

The star topology of the NeLH is illustrated in Figure 3.

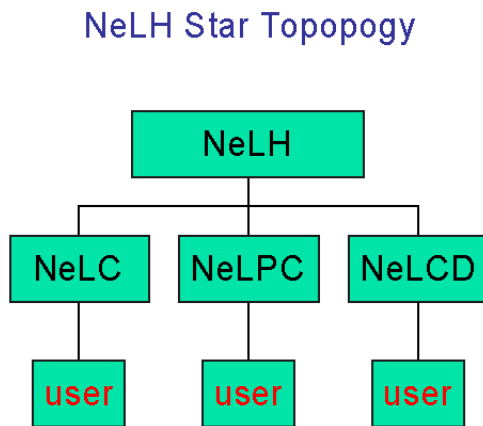


Figure 3: The NeLH Star Topology

4.2. Cross-search

As for the technical issues, NeLCD team is currently investigating the application of Distributed Cooperative Multi-Agent System (CMAS) [18,19] to the NeLH using a star topology. At each VBL, in addition to the Intelligent Agents discussed above, there are also Intelligent Search Agents for Distribution (ISA-D). In addition, there is the Central Intelligence Search Agent (ISA-C) at the top NeLH site.

Intelligent Search Agents for Distribution (ISA-D)

ISA-Ds are present at each VBL and are responsible for ensuring the distributed search.

- They communicate with the particular user ISA to search the local VBL and, secondly, request results from remote VBLs via ISA Central (ISA-C) present in NeLH (discussed below).
- When the results are returned from the ISA-C, the ISA-D is also responsible for combining all results together (local and remote) and presenting them to the user.
- It is activated by a user request to receive all available documents in the NeLH, as the default is to search only the local VBL.

The Central Intelligent Search Agent (ISA-C)

The central NeLH site contains a ISA-C agent responsible for the following tasks:

- keeps up-to-date information of the location and availability of all VBLs (and references to the default ISAs).
- is responsible for combining the results coming from the VBLs ISA agents
- is responsible for filtering out duplications in search results before these are sent back to the user
- and sending the results back the end user, via his VBL ISA-D.
- They could be activated by any ISA-D in VBLs or directly by a user accessing the search facility at the top NeLH site.
- It also communicate with the Negotiation Agent present in NeLH, discussed in the next section.

The architecture is illustrated in Figure 4. Other agents are omitted for reasons of simplicity. The numbers next to communication lines illustrate steps in which the communication takes place.

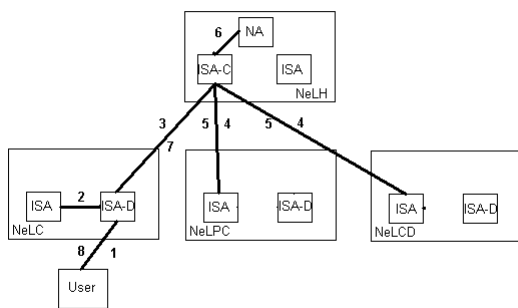


Figure 4: Agents Performing a Distributed Search

Technically, an agent-based query producing an XML output is a robust and flexible solution for all technical platforms used by VBLs. The XML, defined by DTD/schemes based on the discussed DC in Table 1, could be rendered by the receiving VBL to meet the particular VBL requirements on look and feel, thought customisable ISAs, as discussed in [17]. In addition, server-side scripting should be adopted where possible to avoid browser-related problems. Therefore, the output XML sent to the receiving VBL for rendering is transformed into the resulting HTML which is then sent to the client.

4.3. Conflict-resolution

However, a single paper or medical document could come up in several VBLs and could be assigned different a quality tag in each VBL, as their appraisal procedure may vary.

There are two solutions to this problem. Firstly, there is a Negotiating Agent (NA), provided by the NeLH, which will autonomously step in and negotiate a common level of evidence. If agreement cannot be achieved without human intervention, the Negotiation Agent could request the check lists and an additional information from the VBLs involved in order to help the end user to make the final decision on validity of the evidence.

The second approach would be always presenting the user all search results, including those of contradicting quality in order to allow the professional to make the final decision. For these cases, providing the check list (which defines the criteria on which a certain level of evidence was assigned), is essential.

In addition, we are investigating the option of enabling users to customise their search so as they can define whether to authorize the NeLH to perform the negotiation, or whether they prefer to receive all contradicting results and decide themselves.

5. Discussion

The distributed star topology solution was investigated by the NeLCD team. Here we discuss other possible approaches to the library design.

5.1. A Centralized Library

A centralized library would be the best value for money. However, the users vary widely across the libraries (VBLs), and the libraries vary widely in scope and types of content – it is important to maintain the ability of each VBL to respond to the strategic needs of their user base. Therefore, an independence of VBLs in terms of implementing their database structure and search engine, as well as defining the look and feel of the VBL, the user interface and navigation strategy are essential. Also, having a direct control we have more flexibility to adapt and respond to our user base.

5.2. A Peer-to-Peer Topology

As has been discussed above, an alternative solution to the adopted star topology would be a Peer-to-Peer communication. The simplicity of maintaining consistency by the central NeLH server is the main reason for our approach. Therefore, there is just single point of information on VBL location, access, current availability, etc.

Also, all VBLs have to be contacted each time a cross-search is performed to receive all results. As the data are being frequently updated, it cannot be known in advance whether a certain VBL can provide information on a particular subject.

On the other hand, the drawback of this approach is that the NeLH server being as a single point of information is also a single point of failure. Therefore, in case of the NeLH server being unavailable, the cross-search, as discussed above, would not work.

5.3. Search Definitions

We believe that all user queries could be expressed as combinations of logical operators on columns (or full-text search), and therefore, the ODBS type of solutions are not needed. In addition, a number of the libraries have adopted the Lotus Domino database as their implementation platform, therefore, there is no need to provide cross-SQL queries as Lotus database is not relational.

5.4. Sharing User Profiles

Also, user profile could be exchanged among the VBLs to ensure the same functionality regardless of the actual VBL a particular user has logged in. the user could define his preferences regarding resolving conflicts, as discussed above, as well as user interface preferences. This area is also currently being investigated.

6. Project Status

Currently, we have finished our prototype phase and are developing the core library. The distributed search has been investigated by a number of VBL teams and currently we are in the process of finding a mutually convenient technical solution which will be implemented.

As for the cross-search, currently, many VBLs are providing only a search of their local database, some support an explicit call to a server script to perform search in a particular remote VBL, such as VBLs which adopted standardized technical solution called the NeLH Toolkit [20]. However, a coordinated cross-search across all VBLs is still in the design and development phase.

6.1. Implementation Issues

The prototype of the NeLCD library (<http://www.nelcd.co.uk>) has been built using CGI scripts to implement the agent functionality. Also, many other VBLs have adopted proprietary solutions. Currently we are porting the system to Lotus Domino R5 platform, which was chosen as the common environment for all VBLs in the NeLH.

The documents available at the prototype site are in the editorial process now. Therefore, the quality tags and check lists are not available yet.

Also, as some VBL teams are not fortunate in having technical expertise available, a toolkit platform has been developed. This resource is essential for some teams, however, as has been said above, as a great number of libraries adopted a proprietary solution. Therefore, the interoperability issue is crucial to ensure data exchange and distributed search among all VBLs.

As has been said above, the physical distribution of VBLs and full control over technical development of each library is important to ensure flexibility and to respond to specific needs of each VBL users. However, in addition to a communication overhead, there are indications that this solution is not long-term viable for funding reasons. Therefore, it is a matter of discussion whether this project will realize the investigated distributed approach or whether less flexible solution would be the case.

7. Related Work

There are a number related digital libraries providing a collection of cross-searchable documents in the Internet. The Z39.50 [21] standard specifies an abstract information system with a rich set of facilities for searching, retrieving records, browsing term lists, etc. At the server side, this abstract system is mapped onto the interface of whatever specific database management system is being used. The client application is unaware of the implementation details of the software hiding behind the network interface, and it can access any type of database through the same, well-defined network protocol. On the client side, the abstract information system is mapped back onto an interface which can be tailored to the unique requirements of each user. This provides a well founded universal solution, conceptually similar to the one in NeLH, however, the NeLH is a proprietary database which does not aim to provide universality. The same is the case for the general SDLIP communication protocol [22].

A similar approach, looking at a tree hierarchical topology for communicating agents was investigated by Kostkova as the MAGNET Architecture [23].

Also, commercial Web publishing products, such as developed by Interwoven [24], do not provide the

additional autonomous functionality required by the NeLH document quality review and appraisal process, as discussed in [17].

8. Conclusion

NeLH project is developing a distributed digital library providing an information gateway to the best available evidence-based knowledge, enhanced with quality tags, to a wide spectrum of users. In this paper we have introduced the project and discussed the star topology enabling a distributed search across all libraries in the NeLH.

9. References

-
- [1] Gray, J.A.M., S de Lusignan. *National electronic Library for Health (NeLH)*. BMJ, **319**, pp. 1476-1479, (1999).
- [2] The National electronic Library for Health. <http://www.nelh.nhs.uk>
- [3] Medline. <http://www.medline.com>
- [4] M.N. Kamel Boulos., A.V. Roudsari., C. Gordon, J.A.M. Gray. *The Use of Quality Benchmarking in Assessing Web Resources for the Dermatology VBL of the NeLH*, Journal of Medical Internet Research **2001**; 3(1): e5, (2000).
- [5] N. J. Davies, R. Weeks and M. C. Revett. *Information Agents for the World Wide Web*. BT Technical Journal, **14**: 4, (1996).
- [6] J.R. Weinberg, J. Mani-Saada, K. Smith. *The National electronic Library for Communicable Disease (NeLCD)*. Poster. FIS Conference, UK, (2000).
- [7] Public Health Laboratory Service. <http://www.phls.co.uk>
- [8] Cochrane Database. <http://www.cochrane.co.uk>
- [9] C.W. Mulrow. *The medical review article: state of the science*. Ann. Intern. Med., **106**, pp 485-8, (1987).
- [10] A.D. Oxman, G. H. Guyatt. *Guidelines for reading literature reviews*. Can. Med. Assoc. Journal, **138**, pp 697-703.
- [11] NeLH Team. *The Knowledge and Know-how Platform*. Internal document (1999).
- [12] Health Technology Board for Scotland. *Guidance for manufacturers on submission of evidence to Health Technology Assessments*, <http://www.htbs.co.uk/docs/pdf/Consultation%20draft.pdf>, (2002)
- [13] The NeLH Architecture Paper. http://www.nhsia.nhs.uk/nelh/background/arch_index.asp
- [14] P. Kostkova, J. Mani-Saada, J. R. Weinberg. "Do you read the literature properly, and if you do how can others benefit from it?". To appear at the PHLS Conference 2002, Warwick, UK, (2002).
- [15] M.N. Kamel Boulos. *A Preparatory Study on the Dermatology Virtual Branch Library of the NeLH* (MSc Thesis in Medical Informatics), GKT School of Medicine, KCL, University of London, UK, (2000)
- [16] S. Green and F. Somers. *Software agents: a Review*. Intelligent Agents Group, Computer Science Department, Trinity College Dublin and Broadcom Eireann Research Ltd., http://www.cs.tcl.ie/research_groups/aig/iag/iag.html
- [17] P. Kostkova, J. Mani-Saada, J. Weinberg. *Agent-based Up-to-date Data Management in National electronic Library for Communicable Disease*. To appear in the Proceedings of Agents applied in Health Care, 15th European Conference on Artificial Intelligence, ECAI-2002, Lyon, France, July 2002.
- [18] K. Sycara. *Multi-agent Infrastructure, Agent Discovery, Middle Agents for Web Services and Interoperation*. M. Luck et. Al. (Eds.): ACAI 2001, LNAI 2086, 17-49, Springer-Verlag Berlin Heidelberg (2001).
- [19] M. Genesereth and S. Ketchpel: *Software Agents*, Communications of the ACM, **37**, No 7, pp. 48-53, (1994)
- [20] <http://www.nelh.shef.ac.uk/nelh/front.nsf/LL?openform>
- [21] http://www.biblio-tech.com/html/z39_50.html
- [22] <http://www.diglib.stanford.edu/~testbed/doc2/SDLIP/>
- [23] P. Kostkova, J. A. McCann. *Inter-federation Communication in the MAGNET Architecture*. In the Proceedings of the Grace Hopper Celebration of Women in Computing Conference 2000, Cape Cod, Massachusetts, USA, (2000).
- [24] http://www.interwoven.com/products/content_management/key.html