User Customisation of Agent Profiles in the National electronic Library for Communicable Disease

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

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

Gemma Madle
Institute of Health Sciences
The City University
London, UK
G.C.Madle@city.ac.uk

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

ABSTRACT

The Internet provides overwhelming amount of medical information. However, healthcare professionals often cannot find the information they need when they need it and if they do the quality may be uncertain. A new Internet digital library, the National electronic Library for Communicable Disease (NeLCD), is addressing this issue by providing a single-entry portal to evidence-based information on treatment, investigation and prevention of communicable disease. Autonomous Intelligent Agents are essential for the development and runtime of the NeLCD library as they perform autonomously a number of tasks related to the search, assist humans in information publishing, the document review process and data exchange.

In this paper, we present an application of Intelligent Agents in user profiling and customisation. In particular, they allow users to personalise the search, modify the input controlled vocabulary and customize the search results to better meet their needs. In addition, they can autonomously alert users about new postings according to their interests. Profiling of Intelligent Search Agents (ISA) and Pro-active Alert Agents (PAA) allows extensive customisation of the library according to user's personal preferences, professional background and medical specialty.

General Terms

Algorithms, Theory. Digital Libraries

Keywords

Multi-agent Systems (MAS) in Healthcare, Digital Libraries, Autonomous Alert Agents, Intelligent Search Agents, User Customisation, Agent Profiles

1 INTRODUCTION

Despite the fact that healthcare specialists are overwhelmed by the volume of medical information available on the Internet, they often cannot find what they want when and where they need it [1]. This paper focuses on agent-based customisation in the National electronic Library for Communicable Disease (NeLCD) [2], one of the Specialist Libraries (SLs) forming the NeLH.

Digital libraries are one of the key applications for Intelligent Agents on the WWW, where they can perform information retrieval and assist in searching information in the Internet [3]. In addition, agents autonomous behaviour enables them to assist in various search-related processes on users behalf, however, without human intervention. In NeLCD, agents are used to perform a number of tasks, ranging from tracking documents during the review process, through customized search functionality to presenting users interface according to their profile.

The search definition is one of the key parts of any digital library. In NeLCD, we have focused on providing user input by controlled vocabulary based on a pruned MESH tree (NeLH ontology) but allowing users to set their Intelligent Search Agents (ISAs) to customize the pruning process according to their professional specialty, modify the search input and search results GUI. In addition, NeLCD implements customised Pro-active Alert Agents (PAAs) notifying users when a new document of their interest has been inserted into the library.

This paper starts by giving a brief overview of the NeLH project (section 2), and then focuses on the structure of the NeLCD specialist library in section 3. Subsection 3.2 discusses Autonomous Intelligent Agents supporting intelligent searching, and managing the document review process in the NeLCD. In section 4, we present the customisation and user profiling. In section 5 we discuss our work in progress, in particular demonstrating the usage of multi-agent systems (MAS) in information exchange within the scope of the entire NeLH project. Finally, in section 6 we discuss the project status, in section 7 the related work, and in section 8 we conclude.

2 NELH BACKGROUND

The three major goals of the NeLH are:

- 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 (http://www.phls.co.uk), Cochrane database (www.cochrane.co.uk), NHS Centre for Reviews and Dissemination, Effective Health Care Bulletins, British National Formulary, CDC, BMJ and others.

However, the quality, reliability and "non-biased-ness" of provided information significantly vary. For example, studies of Mulrow [4], Oxman and Guyatt [5] 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.

The key approach supported by the NeLH framework is evidence-based healthcare that aims to clearly identify the level of evidence of a study or recommendation.

2.1 Structure of the library

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

Our team is responsible for development of the NeLCD [12], the Communicable Disease branch of the NeLH. 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.

2.2 NeLH Ontology

The main aim of the NeLH is to store evidence-based medical information on clinical and non-clinical subjects, medical teaching material, lectures and slides.

However, there is no common agreement on ontology, nor 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: MESH, CTLV3/SNOMED and ICD10. This is not only a UK but an international issue. The NeLCD team is involved in the EU Agentcities initiative (www.agentcities.net), in particular the Health care working group (www.agentcities.org/EUNET/Activities/WG/Healthcare) looking at, among others, existing ontologies in participating EU countries and finding options for common agreements.

However, MESH (www.nlm.nih.gov/mesh/2002/index.html) – Medical Subject Headings was chosen by the NeLH team as the underlying medical ontology for indexing documents in the library. As it is the most commonly used among medical librarians in the UK and is the standard used for indexing in Medline. In the Dublin Core (DC) definition, keywords appear as "Subject". Agent support for user customisation and personalization of the indexing scheme is discussed in greater detain in the technical section 4. There have been indications that a modified SNOWMED ontology may be adopted in the future stages of the NeLH development, however, this is currently a matter of discussions.

In order to support a user-customisable 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 cataloguing 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 [6] describing a preparatory study of metadata, RDF and problem-knowledge coupling for the NeLH.

Every document in the NeLH will be described by the following fields – called "electronic catalogue card".

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

Table 1: The Electronic Catalogue Card

3 STRUCTURE OF THE NELCD

In this section, we will look internal data representation in the NeLCD and agents deployed in the library.

3.1 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.

Only pre-processing and quality-tagging of available information before incorporating it into NeLCD could ensure the required

quality of the site. However, 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).
The Timestamp attached to every document in NeLCD will consist of three data items: Publication date, Posting date (to the NeLCD database) and Expiry date (date when document is reviewed by the NeLCD team, usually 1 year after posting and annually there after).
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" (
Table **1**) has passed which is performed by the Reactive Expiry Agents.
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. The quality tag consists of:

- Level of Evidence (meta-analysis, Randomised Control Trial, Clinical Trial, Cohort Study and other)
- Reviewers Assessment critically summarizing the paper, and
- Checklist, which answers brief questions about the methodological issues, level of evidence, potential biases and applicability of the results (http://www.staff.city.ac.uk/~jane/check.doc)

The resultant quality tag and a signature of the particular society are attached to the document and made available through the NeLCD.

3.2 Agents 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 which is discussed in detail in section 4. The latter, Reactive Review Agents and Reactive Expiry Agents, are in charge of the library review process. The NeLCD MAS is illustrated in Figure 1. Here, we give a brief overview of their functionality, the full description is discussed in [12].

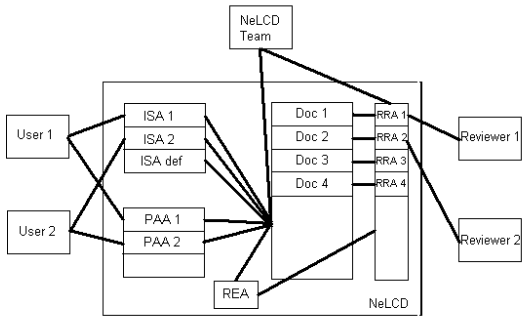

Figure 1: Agents in NeLCD

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 RCT. Also, ISA could be personalized to adapt the keywords for GUI according to user profile – this is discussed in the following section on customisation.
Pro-active Alert agents (PAA) are performing tasks on users-behalf, such as alerts according to user interests. Detailed description of these agents is provided in the next section.
Reactive Review Agents (RRA) and the Reactive Expiry Agent (REA) are responsible for autonomous management of the review process, the former ensure that each document is assigned a reviewer and that the Reviewers Assessment (RA), see the "electronic card" for description, is obtained and displayed on the NeLCD site. While the latter is ensuring that each document is updated when its "Expiry Date" has approached.

4 AGENT-BASED PERSONALIZATION AND CUSTOMISATION IN NELCD

In this section, we present how users can customize the search, personalize the controlled vocabulary used for keyword-based search data input and discuss in greater detail the remaining agents: ISAs and PAAs.

4.1 Indexing and Search Definitions

As has been introduced above, the library metadata structure is the Dublin Core (
Table 1) and the chosen ontology is the MESH indexing scheme. Primarily, MESH terms are used for indexing the documents in the library (DC field "Subject").
The second area where we use MESH in NeLCD is presenting a pruned MESH tree as a controlled vocabulary for user search input. The pruning was performed to reduce the number of the keywords and reflect the needs of the communicable disease domain. Therefore, MESH sub-trees B "Organism", C "Diseases" and parts of the sub-tree D "Chemicals and Drugs" were originally selected; the number of levels included was chosen according to the frequency of each particular term. Typically, the pruning process included terms up to the 3rd or 4th MESH tree level. Figure 2 illustrates a small sample of the pruned sub-tree "C'.

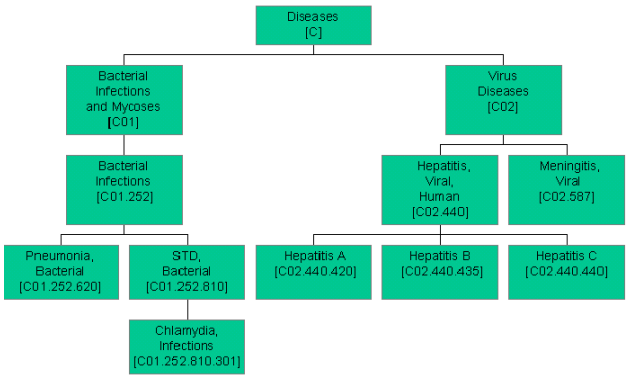

Figure 2: Example of Pruning the MESH tree

Users can also input free text for search themselves, however, as these typically vulnerable for spelling mistakes, language differences, etc, therefore, the preferred way for search data entry is the controlled vocabulary based on MESH. The NeLH is looking at providing a medical thesaurus for the VBLs to support the free-text data entry, however, selecting keywords from pull-down controlled vocabulary lists gives better performance and is more user-friendly solution.
In addition, we believe that all user queries could be expressed as combinations of logical operators on MESH keywords and other DC fields (e.g., user can define to search for randomised control trial on TB and meningitis), and, therefore, DBMS type of solutions are not needed.
In our future research, in collaboration with the other VBL teams and the NeLH, we will be investigating a concept-based approach to the free-text search input which, based on MESH as the underlying ontology, will be able to generate concept-related queries and context-relevant terms and map them onto the MESH terminology.

4.2 User Profiles and Customisation

Firstly, users come from different professional backgrounds and specialties (microbiologists, clinicians, GPs, nurses, etc.) which determine their medical interests. In addition to their specialty, they may have particular interests in treatment or investigation of a particular disease or a group of diseases. The agent system allows them to set up profiles for their ISA and PAA, based on the MESH index scheme, to customize the functionality of their ISA and PAA. ISAs' profile can customize the pruning process to tailor the search input controlled vocabulary to the user interests; PAA's profile can define a set of keywords to ensure that the user is alerted by the PAA when a relevant entry has been inserted. This is discussed in detail in the next sessions.

In addition, users can customize the GUI in terms of specifying their preferred look of their NeLCD search pages, e.g., changing defaults for search to set the preferred DC field values (e.g., search for RCT, displayed no later than 1 week).
Also, they can modify the search results page to set the number of items to be displayed. The default setting displays all found items.

4.3 Intelligent Search Agents

The primary goal of the library to provide a single portal for searching for an up-to-date medical information. Intelligent Search Agents (ISA) are used to provide the search-related functionality.

Tasks

- Presenting users with a user-friendly adaptive interface to define a query for searching medical information in the library.
- Performing the searching functionality according to the given criteria.
- Displaying the search results according to user given criteria.

Number

There is an ISA for each registered user to allow the customisation of the interface and the search. In addition, as registration is not compulsory, there is a default ISA providing standard search and presenting default interface to non-registered users.

Interaction and Coordination

Each ISAs interacts with one user – that is, a customized ISA dedicated to a particular user provides the customized functionality, and a number of default ISAs serve non-registered users.
In addition, they also communicate with the Knowledge Sources to perform the search and return the results back to the user.
When a user logs in the site the dedicated ISA is activated and the user interface and search is customized accordingly. A new default ISA is created when a new user logs in.

User Customisation of ISAs

There are four features that the ISA profile consists of:

1. Firstly, as has been introduced above, users can customize the search entry page by setting their *criteria for the pruning process* to indicate which sub-trees (starting with the default pruning discussed in section 4.1) are not interested in and which they want to be listed in deeper levels. ISAs then customise the controlled vocabulary.
2. 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 Randomised Control Trials, or Case Studies
3. ISA's profile can also customise the *look of the search page*, e.g., colour coding and the page layout. We are currently investigating this feature.
4. Finally, all search results are returned to users by their ISAs by default. Users can set *how many entries* they want to be displayed and indicate if and *how the remaining results should be displayed*. In particular, this is a useful feature for those searching the site using a mobile device.

4.4 Pro-active Alert Agents

Pro-active Alert agents (PAA) are performing tasks on users-behalf.

Tasks

- Users can instruct PAAs, to perform given tasks on their behalf autonomously or alert them accordingly. In particular, the PAA is monitoring the library and informing users about new postings related to user -specific issues.

Number

There is a PAA for each registered user to perform the alert functionality. Non-registered users have no PAA agents.

Interaction and Coordination

PAAs interact with users – one PAA per registered user. In addition, they communicate with the Knowledge Sources to check changes in the database interested to each particular user.
PAA act autonomously with a frequency given by the particular user – e.g., every day, once a week, etc. Users can set the PAA to alert about several issues, each with different frequency. PAA is in control of its activation according to the settings.

User Customisation of PAAs

The PAA's profile is set by users to indicate any DC field values the user is interested in. The profile is based on keywords from the MESH tree, however, any fields value could be set (e.g., all documents on HIV inserted or updated in last month). A typical action performed by the PAA is an email message sent to the user with appropriate links when a new relevant document has been inserted.

From these agent descriptions is it clear that a particular user ISA profile differs from his PAA profile. In addition, we are currently investigating a framework for adaptive ISA agents learning user profile dynamically based on the user common search terms, however, for performance demands of the pruning process and the alert agents, users would have to approve the dynamic profile before the agents start using it.

5 DISCUSSION AND FUTURE WORK

In this section we outline areas for future research.

5.1 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.

5.2 Real-time Pro-active Alert Agents

The NeLCD is not just an agent-driven digital library, but servers also as a platform for building an online community around communicable disease. We are currently looking at building chat rooms and moderated discussions to allow professionals to discuss current medical issues online.
The library is open for public, however, the chat rooms will be accessible only for registered professional users.
We are currently investigating a new concert of PAAs - Real Time Pro-active Alert Agents (RT-PAA) that will monitor the chat rooms and alert the user when a topic of his interest is currently being discussed, or if his colleague has joined a particular discussion.

6 PROJECT STATUS

Currently, we have finished our prototype phase and are developing the core library. The major professional societies involved in the practice of communicable disease have indicated their support to the project and an initial model of collaboration have been set up. Also, the first "Reviewers Assessments" have been provided by society members.

6.1 Implementation Issues

The prototype of the NeLCD library (http://www.nelcd.co.uk) has been built using CGI scripts to implement the basic agent functionality. The top NeLH site could be found at http://www.nelh.nhs.uk.

Currently we are porting the system to Lotus Domino R5 platform, which was chosen as the common environment for all VBLs in the NeLH. Lotus Domino built-in agents support is used for implementing the agent functionality discussed in this paper.

The prototype of the NeLCD library was populated with a core evidence-based documents with relevance to the infections diseases, syndromes and presentations. Documents in NeLCD are represented in XML (relevant DTD is defined for validation purposes). Common DTD across NeLH will allow extensive document exchange and cross-VBL search performed by Information Agents. The separation of the content from style allows flexible manipulation with data, easy modifications of the display format, as well as object-based data representation suitable for data search and document exchange. The documents available at the prototype site are in the editorial process now and the first Reviewers Assessments have been received.

Also, as some VBL teams are not fortunate in having technical expertise available, a toolkit platform has been developed (http://www.nelh.shef.ac.uk/nelh/front.nsf/LL?openform). This resource is essential for some teams, however, 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.

Ethical aspects of any Internet-based healthcare project need to be addressed. In the case of NeLCD, no personal patient data are gathered or processed. User profiles collected by Information Agents are kept and used with the owner consent. In addition, copyright aspects are not an issue either as the library provides links to documents which are already in the public domain [7].

7 RELATED WORK

Autonomous agents providing a wide range of functionality in health care applications have been an interesting area of research in recent years in academia and industry.
For example, agent project by Honeywell is investigating applications of autonomous agents in elderly patients nursing [8], agentcities-funded (www.agentcities.net) healthcare project is looking at implementing MAS for negotiating patients visits to specialists according to his or her condition and physical location [9]. Agents-assisted recommendation for screening of cancer patients and other projects were investigated by Cancer Research Fund, UK [10]. Web crawling agents discovering biomedical information were also explored [11]. However, the NeLCD seems to be the only project using agent technology building a dynamic medical digital library.
There are a number related digital libraries providing a collection of cross-searchable documents in the Internet. The Z39.50 (www.biblio-tech.com/html/z39_50.html) standard specifies an abstract information system with a rich set of facilities for searching, retrieving records, browsing term lists, etc. 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 (www.diglib.stanford.edu/~testbed/doc2/SDLIP).
Also, commercial Web publishing products, such as developed by Interwoven do not provide the additional autonomous functionality required by the NeLH document quality review and appraisal process, as discussed in [12].

8 CONCLUSION

NeLCD project is developing a digital library providing the best available evidence-based knowledge to a wide spectrum of users: clinical experts, public health, general practitioners and general public. In this paper we have discussed how NeLCD fits into the NeLH project and focused on the application of Intelligent Agents in user profiling and customisation.

Firstly, user search input and search results definitions can be modified. Pruned MESH tree reflecting the needs of communicable disease domain serves as the default controlled vocabulary presented to users for search input. The tree could be

customized by setting the ISAs to personalise the pruning process to better meet the user needs. Secondly, PAAs can alert users according to his or her interests.

In addition, we have outlined a CMAS solution to the distributed search and the options for sharing user profiles among VBLs.

9 REFERENCES

[1] J.A. Muir Gray, S de Lustignan. *National electronic Library for Health (NeLH)*. BMJ, **319**, 1476-1479, (1999).

[2] P. Kostkova, J. Mani-Saada, J. R. Weinberg. *"Do you read the literature properly, and if you do how can others benefit from it?"*. PHLS Annual Scientific Conference 2002, Warwick, UK, (2002).

[3] N. J. Davies, R. Weeks and M. C. Revett. *Information Agents for the World Wide Web*. BT Technical Journal, **14: 4**, (1996).

[4] C.W. Mulrow. *The medical review article: state of the science.* Ann. Intern. Med., **106**, pp 485-8, (1987).

[5] A.D. Oxman, G. H. Guyatt. *Guidelines for reading literature reviews*. Can. Med. Assoc. Journal, **138**, pp 697-703, 1988.

[6] 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)

[7] P. Kostkova, J. Mani-Saada, J.R. Weinberg. *Ethical Issues in Health Care on the Internet: "Much ado about nothing?",* In the Proceedings of the Nordic interdisciplinary conference and workshop at NTNU "Making Common Ground: Methodological and Ethical issues in doing Internet research", Trondheim, Norway, June 2002.

[8] K. Z. Haigh, J. Phelps, C.W. Geib *An Open Agent Architecture for Assisting Elder Independence.* In the Proceedings of the AAMAS 2002, Bologna, Italy, July 2002.

[9] A. Moreno, D. Isern. *A first step towards providing health-care agent-based services to mobile users.* In the Poster session of AAMAS 2002, Bologna, Italy, July 2002.

[10] J. Fox. *Understanding intelligent agents: analysis and synthesis* In the Proceedings of Agents applied in Health Care, 15th European Conference on Artificial Intelligence, ECAI-2002, Lyon, France, July 2002.

[11] P. Srinivasan et. al. *Web Crawling Agents for Retrieving Biomedical Information*. BIXMAS 2002, Bologna, Italy, July 2002.

[12] P. Kostkova, J. Mani-Saada, J. Weinberg. *Agent-based Up-to-date Data Management in National electronic Library for Communicable Disease*. In the Proceedings of Agents applied in Health Care, 15th European Conference on Artificial Intelligence, ECAI-2002, Lyon, France, July 2002.