Medical Informatics Europe '99
P. Kokol et al. (Eds.)
IOS Press, 1999

CISMeF: Cataloque and Index of French speaking health resources

SJ. Darmoni (a), JP. Leroy (a), F. Baudic (b), M. Douyère (a), J. Piot (b), B. Thirion (b)
(a) Computer and Networks Department, (b) Medical Library,
Rouen University Hospital, 1 rue de Germont 76031 Rouen Cedex

Abstract. In 1999, the Internet has become a major source of health information. The objective of CISMeF is to catalogue and index the main French-speaking sites and documents concerning health. Currently, the number of resources already totalled over 6,100 with a mean of 75 new sites each week. CISMeF contains a thematic index, including medical specialities and an alphabetic index. CISMeF uses two standard tools for organising information: the MeSH (Medical Subject Heading) thesaurus from the Medline bibliographic database (National Library of Medicine) and the Dublin Core metadata format. A brief description of the site is systematically added. CISMeF respects the Net Scoring, criteria to assess the quality of health information on the Internet. The CISMeF project fulfils a valuable tool for the French-speaking health community: 2,500 machines visit the Web site each working day.

MeSH key-words: Abstracting and indexing; Cataloging; Internet; Vocabulary controlled

1. Introduction

In 1999, the Internet can be considered as a major source of scientific and health information. For the health professional, to find the accurate information on the Internet is not so easy; therefore, there is a profusion of directories such as Yahoo and search engines such as Altavista. In a previous study [1], we demonstrated that manually catalogues vs. search engines were less sensitive but far more specific.

The objective of CISMeF is to assist health professional during his/her search of electronic information and knowledge available on the Internet. CISMeF is a project initiated by the Rouen University Hospital (RUH). Its Universal Resource Locator (URL) is http://www.chu-rouen.fr/cismef. CISMeF began in February 1995 with the creation of the RUH's Web site. The scope of CISMeF covers healthcare disciplines and medical sciences.

2. Material and methods

– *Hardware and software.* CISMeF was implemented in February 1995 on a Sun running Sun Unix. CISMeF is a Web Site using the NCSA freeware http server. The Gestats and Webtrends programs evaluate the use of the Web page after excluding requests by our Hospital Information System computers and image files.

– *Standards.* CISMeF is entirely based on static HTML. CISMeF uses two standard tools for organising information: the MeSH (Medical Subject Heading) thesaurus from the Medline bibliographic database (US National Library of Medicine) and the Dublin Core metadata format [2].

- *Methodology and realisation of the catalogue:*

Since its creation, CISMeF has only catalogued and indexed French-speaking health resources, independently of its origin. The CISMeF method entails a five-fold process: resource collection, filtering, description, classification, and index.

Resource collection: A deputy medical librarian (J.P.) performs the resource collection on French-speaking health resources on a daily basis. She checks French-speaking directories and search engines, especially their "what's new" pages. A total of 420 health webmasters (7% of the sites included in CISMeF) have sent us an Email or a specific form to be indexed in CISMeF. Indexing priority is given to Internet sites of institutions and scientific societies. Resources include sites and high quality documents, issued especially from evidence-based medicine, practice guidelines, consensus development conference, teaching material, and technical report.

CISMeF respects the Net Scoring [3], which consists of 48 criteria to assess the quality of health information on the Internet. These criteria are grouped in eight categories: credibility, content, links, design, interactivity, quantitative aspects, ethics, and accessibility. Some of the criteria are inspired by a US white paper [4]. The description of a site should permit the evaluation of the quality of its content. Some sites and documents are not introduced in CISMeF, because they do not respect basic particularly ethical criteria.

Catalogue and index: Cataloguing a site is necessary because it helps to estimate, in advance, the type of information, which saves time for the end-user, and to evaluate its content. Resources included in CISMeF are described by the following elements from the Dublin core project [http://purl.org/DC/about/element_set.htm]: title, author or creator, subject and keywords, description, publisher, date, resource type, format, identifier, and language. We also include the following elements from the Dublin core metadata format in the metadata of CISMeF's HTML documents: language, and keywords and their synonyms in English and in French. We have organised our catalogue with the MeSH thesaurus, which is precise, rigorous and annually updated. We also use the French translation of this thesaurus, performed by the French Medlars Center, the National Institute for Health and Medical Research (INSERM).

In CISMeF, each keyword is "de facto" a MeSH Major Topic. MeSH subheading permits a focus on a sub-field of a MeSH term, e.g. chloride/toxicity. We also use a French translation of the MeSH subheadings, which is less systematically used in CISMeF compared to Medline. CISMeF resource type is a generalisation of the publication type of Medline. We have added types which are specific of the resources available on the Internet, such as association, patient information, community networks (see http://www.chu-rouen.fr/documed/typeressource.html). Resource type describes the nature of the resource and MeSH describes the subject of the resource.

Example of a description of a document indexed in CISMeF:

> Adolescent pregnancy [http://www.cps.ca/english/statements/AM/am94-02.htm]. Adolescent Medicine Committee, Canadian Paediatric Society (CPS). In : Canadian Journal of Paediatrics 1994; 1(2): 58-60 [prevention, diagnosis, management, counselling, bibliography]. -Ca
> *keywords : adolescence ; pregnancy in adolescence*
> *resource type : clinical practice guideline*

Structure of the CISMeF catalogue: CISMeF contains a thematic index, including medical specialities and an alphabetic index. A brief description of any site indexed in CISMeF is systematically added. To index the health resources CISMeF uses five levels of hierarchy: (1) CISMeF "meta-term" (2) MeSH category (3) MeSH keyword (4) MeSH subheading (5) CISMeF resource type. The levels 1 and 5 are specific of CISMeF. The levels 2, 3 and 4 are using the Medline database structure.

The CISMeF "meta-term" is generally a medical speciality or a biological science, e.g., cardiology or bacteriology. For example on the cardiology page, the sites of general interest about this speciality are indexed and described, followed by a list of starting points of related

categories and others associated MeSH terms. In CISMeF, each MeSH term (level 3 of the CISMeF structure) corresponds to a HTML document, which is organised first with MeSH subheading (level 4), then for each subheading, the resource types (level 5). The alphabetic index uses the MeSH terms in English and their French translation, which permits bilingual search. Alphabetic and thematic indices are available, and also a general index for permuted utilisation. Currently, CISMeF contains indexed 6,118 resources with 1,444 MeSH terms. A mean of 75 sites and documents are indexed each week and approximately 32 new MeSH terms are added in the same time.

3. Results

CISMeF is efficient an end-user-friendly solution to find French-speaking worldwide health resources on the Internet. Two-thirds of these sites and documents are located in France, 20% are from Canada, in particular the Quebec Province, 5% from Switzerland, 4% from Africa and 3% from Belgium. This Web site is principally and initially oriented for the health professional, although the general public may also have access to it. Many sites are devoted to both. There are no HTML documents with restricted access in the CISMeF Web site. Thus our traditional "end-users" are now not only healthcare practitioners but also patients, their families and health consumers. One deputy medical librarian (J.P.) performs the resource collection and the information watch. Two deputy medical librarians (F.B. and M.D.) describe and index resources. The chief medical librarian (B.T.) is a 'super-indexer' in charge of checking the indexing. There is a 30-minute meeting daily with the medical informaticians (S.J.D. and/or J-P.L.) for double-checking. Since February 1995, some new features have been added to optimise the navigability and the access to the information for the end-user: (a) use of an internal search engine (full-text search), (b) a general index, and (c) a "what's new" page to easily display the newly indexed sites on a weekly basis. Since January 1997 it also includes an archive of the what's new pages. Two guides to use CISMeF are also on line, one for basic search and one for advanced search. CISMeF is accessible by the lowest common denominator of current browser technology. Since December 1996, CISMeF has an Editorial Board, following the policy of the hospital managers to control its development, its quality, its validity, its reliability, its consistency and its coherence. The editorial board peer reviews each document and site included in CISMeF. If necessary, we use external peer reviewing.

Use patterns of the Web site: Analysis of a representative period, the month of March 1999, showed that every working day approximately 2,500 machines visited our site (excluding ours). During the entire month, users from 68,601 computers made 280,195 requests for HTML documents originating from 114 different countries (33.58% from France, 23.48% from USA, and 8.03% from Canada). These statistics underestimate the real figures due to the practice of file cacheing. The CISMeF "Web impact factor" [5] (number of sites, which have at least one hyperlink to our site) was over 550. Over 110 press articles released information about our Web site. In March 1998, CISMeF has obtained the label "Experimentation of public interest" by an inter-ministry committee (procedure " Information Highways").

4. Discussion

The Internet facilitates the communication among the health professionals and with the general public, and also improves the information access. We distinguish several tools in the retrieval of health information on the Internet: - *level 1*: search engine, generalist or more specialised ones, such as MedHunt; - *level 2*: catalogue and index without thesaurus, such as and Medical Matrix, US; - *level 3*: catalogue and index with thesaurus, such as the UMLS (Unified Medical Language System) metathesaurus and MeSH thesaurus. The latter thesaurus is used in the following health catalogues, MedWebPlus, CliniWeb [6], Oregon Health Sciences

University-US, and DDRT (Diseases, Disorders and Related Topics, Medical Library and Medical Information Center, Karolinska Institute, Stockholm, Sweden) ; - *level 4*: catalogue and index with thesaurus, metadata, and description of sites. CISMeF and two other catalogues have now reached this level: OMNI (Organising Medical Networked Information - UK) [7] and HON (Health on the Net - Switzerland) [8]. OMNI and MedWebPlus, Us are also using the UMLS metathesaurus to provide a conceptual network to the subject headings. OMNI, HON, and CliniWeb have also developed a structured database (dynamic HTML) which permits better searches. This improvement will be developed for CISMeF in 1999. OMNI and CISMeF are using the Dublin Core metadata format, which is expected to become the dominant metadata format for Internet resource description.

One main objective of CISMeF is to promote best medical practice and teaching. Therefore, we index high-quality documents available on the Internet on a priority basis. Further challenges that CISMeF needs to address in the next months are to expand sites and high-quality documents, especially patient information, and to collaborate more closely with similar services, particularly in Europe (DDRT, HON and OMNI).

Acknowledgements

CISMeF was supported in part by the grant n° 1998 06 016 from the Agence Universitaire de la Francophonie in the program "French-speaking Virtual University". The authors thank Richard Medeiros for his advice in the editing of this manuscript.

References

[1] SJ Darmoni, B. Thirion,. Indexing the Web ? A comparative study of three medical Web servers on the Internet: Cliniweb ,"Diseases, Disorders and Related Topics", *Omni. Mednet 96, European Congress of the Internet in Medicine* (1996) 5-6.
Available from Internet: <http://www.mednet.org.uk/mednet/dl4.htm>.

[2] S Weibel, H. Juha, DC-5: The Helsinki Metadata Workshop; A Report on the Workshop and Subsequent Developments. *D-Lib Magazine* February (1998)
Available from Internet: <http://www.dlib.org/dlib/february98/02weibel.html>.

[3] Centrale Sante. Net Scoring : criteres de qualite de l'information de sante sur l'Internet; 20 Apr 1998 [Web document, accessed 27 Apr 1999].
Available from Internet: <http://www.chu-rouen.fr/dsii/publi/critqualv2.html>.

[4] J. Ambre, R. Guard, FM Perveiler, J. Renner, H. Rippen, Health Information Technology Institute. Working Draft White Paper: Criteria for Assessing the Quality of Health Information; 20 Apr 1998 [Web document, accessed 27 Apr 1999].
Available from Internet: <http://www.mitretek.org/hiti/showcase/documents/criteria.html>.

[5] P. Ingwersen, The calculation of WEB impact factor, *Journal of Documentation* **54 (2)** (1998) 236-243.

[6] WR. Hersh, KE. Brown, LC. Donohoe, EM Campbell, AE Horacek, CliniWeb: managing clinical information on the World Wide Web, *Journal of the American Medical Informatics Association* **3(4):** (1996) 273-280.

[7] F.Norman, Organising Medical Networks' information (OMNI). *Med. Inf.* **23** (1998) 43-51.

[8] C. Boyer, O. Baujard, V. Baujard, S. Aurel, M. Selby, RD. Appel, Health On the Net automated database of health and medical information, *International Journal of Medical Informatics* **47(1-2)** (1997) 27-29.