

Modelisation of Consumer Health Information in a Quality-Controlled Gateway

Lina F. Soualmia^{a,b}, Stéfan J. Darmoni^{a,b}, Magaly Douyère^a, Benoît Thirion^a

^aComputer & Network Department, Rouen University Hospital, 76031 Rouen, France.

^bPerception, Information, Systems, INSA-Rouen, 76131 Mt St Aignan, France.

Abstract

The amount of health data accessible on the Web is increasing and Internet has become a major source of health information. Many tools and search engines are available but medical information retrieval remains difficult for both the health professional and the patients. In this paper we describe CISMef-patients. It is a sub-part of CISMef, a structured quality-controlled subject gateway. CISMef-patients has been designed for the patients, their families and the general public who are often unfamiliar with the medical domain and the medical vocabulary. The resources included in CISMef-patients are described using the Dublin Core metadata format. To index them, CISMef-patients and CISMef share the same terminology, which 'encapsulates' the MeSH thesaurus with a layer of synonyms. Unlike Medline-plus and Medline, sharing the same terminology allows to a CISMef-patients end-user the possibility to extend his query to the CISMef catalogue (e.g. for searching teaching resources or clinical guidelines).

Keywords:

Internet; Patients; Health Information consumer; Cataloguing; Information Retrieval; Metadata; Controlled Vocabulary; Terminology; MeSH

1. Introduction

Internet is a major source of health information. Many people including professionals, patients and general public, now search health care information on the Web. The access to structured medical information remains difficult when using directories such as Yahoo [<http://www.yahoo.com>] or search engines such as Altavista [<http://www.altavista.com>]. Therefore many tools and applications have been developed for the healthcare professionals and until recently databases such as Medline were available only to experts [1]. The patients and the general public are also interested by health care information and now the importance of the Internet as a source for health information for the layperson is increasingly acknowledged [2, 3] and the growth rate in lay use of Internet health sites is rapid [4]. The information downloaded by the patients must be of good quality and reliable which is not always the case on the Internet [5]. Health information consumers, such as the patients and the general public, cannot assess themselves of the good quality of the information. Several studies [6] have identified other kinds of problems. In medical information retrieval, there is a need of support for the health consumer: when an end-user is searching for information, he is not always familiar with the medical domain and vocabulary. This implies a bad query formulation when he is unaware to express a request, syntactical mistakes when he doesn't know the specific medical term to employ, or more simply some misspelling. Another problem is the readability of the information accessed.

Nevertheless, electronic documents and web sites that provide public access to health information exist, for example Medline-plus [7] [<http://www.medlineplus.gov>] and Patient-Uk [<http://www.patient.co.uk>].

In this context of bad information retrieval, the objective of CISMef [8] [<http://www.chu-rouen.fr/cismef>] (Catalogue and Index of French-speaking resources) is to assist the health professional during the search of electronic information available on the Internet. The CISMef catalogue references high quality information resources. Due to the important quantity of information written for the patients and the general public, the sub-catalogue CISMef-patients [9][<http://www.chu-rouen.fr/patient>] has been designed in 1997. Such as CISMef, it includes only quality-controlled health information resources and those that are dedicated to the general public. The objective of CISMef-patients is a useful information retrieval for the non-professional, avoiding many problems such as knowing the very specific medical vocabulary.

2. Material and Methods

The CISMef catalogue describes and indexes a large number of health information resources (n=11,437). Each catalogue resource is indexed by its container using metadata and by its contents using the terms of the CISMef terminology. Metadata are data concerning the data. They refer to descriptive information about the Web resources [10] and are used to improve information retrieval [11]. In CISMef, the Dublin Core [12](DC) metadata format is used. Each resource is described by eleven of the fifteen elements of DC and for more precision in the resource description, eight elements [8] specific to CISMef were added: *institution*, *city*, *province*, *country*, *target public*, *access type*, *cost* and *sponsorship* of the considered resource. The HIDDEL metadata vocabulary (Health Information Disclosure Description and Evaluation Language [13]) for health information quality is used since November 2002. CISMef-patients and CISMef share the same information model and terminology.

The CISMef terminology 'encapsulates' the MeSH thesaurus [14]. Many other health catalogues are based on the MeSH thesaurus. For example: Health On the Net-Ch [<http://www.hon.ch>], DDRT-Se [<http://www.mic.ki.si/Diseases/>], CliniWeb-US [<http://www.ohsu.edu.cliniweb>], Oregon Health Sciences University-US, Organizing Medical Networked Information Uk [<http://omni.ac.uk>] and Healthinsite-Au [<http://www.healthinsite.gov.au/>]. In the CISMef terminology, in addition to MeSH keywords (n=20,742) and qualifiers (n=83) in its 2002 version, the concepts of metaterms (n=60) and resource types (n=115) were added. A metaterm [8] is a medical speciality or a biological science, e.g. '*cardiology*' or '*bacteriology*', and it is in most cases a MeSH term. The metaterms were introduced to cope with the relative restrictive nature of the MeSH terms when searching '*guidelines in cardiology*' or '*databases in virology*' where *cardiology* and *virology* are metaterms and *guidelines* and *databases* are resource types. The resource types are a generalisation of the publication types of Medline. The keywords, qualifiers and resource types are hierarchically organised. Each metaterm has a semantic link with one or more keywords, qualifiers and resource types. Each term can have a set of synonyms and can belong to several trees. For example the term '*skin tumour*' is associated with the metaterms *dermatology* and *cancerology*.

CISMef and CISMef-patients are quality-controlled gateways such as defined by Koch [15]. The following elements characterise a typical quality-controlled subject gateway and are fulfilled in CISMef:

- Selection and collection development

- Collection management
- Intellectual creation of metadata
- Resource description (a metadata set)
- Resource indexing (with controlled vocabulary system).

In order to include only reliable resources, and to assess the quality of health information on the Internet CISMef uses the main criteria (e.g. source, description, disclosure, last update) of the Net Scoring [16]. There are 49 criteria [http://www.chu-rouen.fr/dsii/publi/netscoring.html] chosen by a consensus of experts. Some resources are not included in the catalogue, because they don't respect basic, particularly ethical criteria. The quality of health information is a key point to consider, specifically for the patients and their families. In the indexing process step, a descriptive annotation (an HTML or XML file, and since November 2002 an RDF file) is created by the librarians and associated to each selected resource, using Dublin Core metadata element set, CISMef metadata and HIDEEL. This allows the indexing of a resource by its informational container. A set of keywords, qualifiers, and resources types according to the CISMef terminology allows the content indexing.

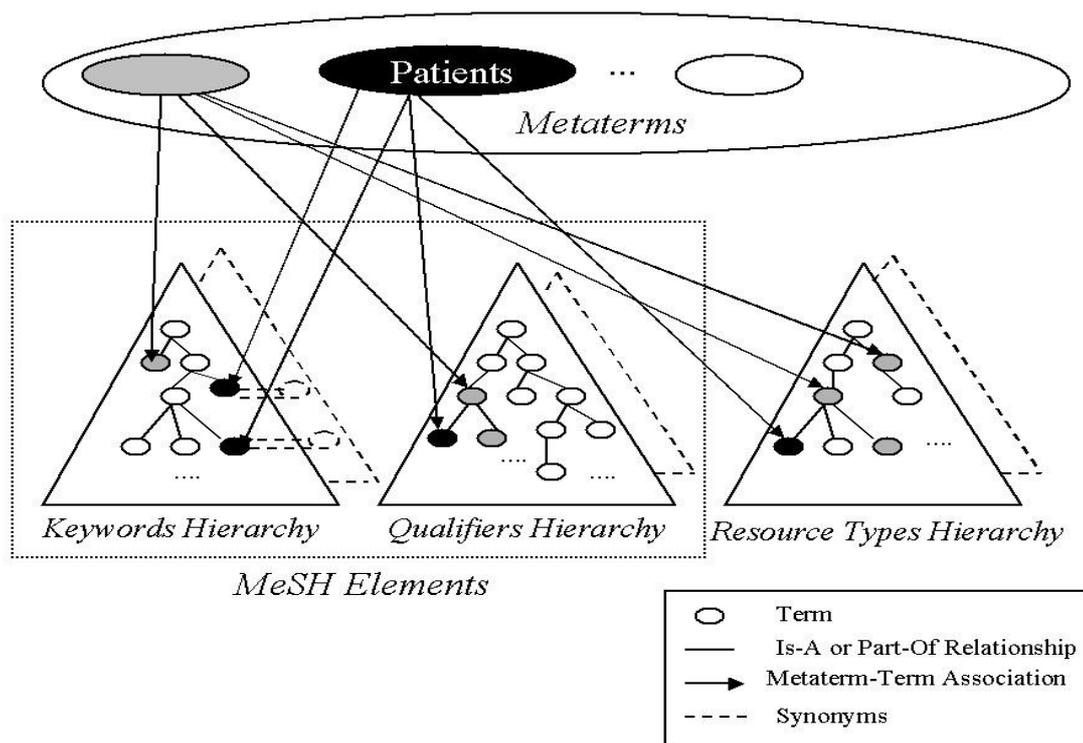


Figure 1 The CISMef terminology structure

3. Results

CISMef-patients [9] developed since 1997, is a sub-catalogue of CISMef [8]. It has been designed to help users such as the patients, their families and the general public in their search for information. The CISMef model and vocabulary is not easy to understand for a non-professional. CISMef-patients can be considered as a specific view on the CISMef terminology. This view corresponds to the metaterm *Patient*, which has its own keywords'

tree structures. To allow a simpler navigation, popularised synonyms were associated to the terms at each level of the model. The synonyms are terms used in the popular language (e.g. '*mania*' is a popular synonym of '*bipolar disorder*') and were determined thanks to a collaboration with patients' associations. The navigation into CISMef-patients can be done through an index of medical specialities (n=34). A general index of all terms used in CISMef-patients is also available (n=343). This avoids the layperson to type its queries. The tree structures are visualised in a simpler way. The CISMef-patients catalogue is composed by information resources, dedicated to the patients and the general public and written by the health professionals, medical institutions and associations of patients.

For searching for information, preformatted queries are applied on Doc'CISMef [17] the search tool associated to CISMef. The preformatted queries are generated automatically when a user clicks on a keyword link. Therefore, the user is neither obliged to know the MeSH keyword, nor the query language used in the Doc'CISMef search tool. The corresponding query generated is the following equation:

Patient [metaterm] AND [keyword] (1)

For example, in the case of clicking on *mania* the generated query is:

Patient [metaterm] AND bipolar disorder [keyword] (2)

It should be noted that it was difficult to the CISMef team to choose the preformatted request. The alternative was between a general query with the metaterm *Patient*, or a more precise query limited to patients' resources types with the following query:

Patient [resource type] AND [keyword] (3)

The patient resource types are the following ones:

- Hot lines
- Associations of Patients
- News Group and discussion list for patients
- Patient and health consumer information

A whole of 1,357 information resources are described and indexed in the CISMef-patients catalogue with the metaterm *Patient*. Another type of access is available via the *life periods*. It is similar to the *life events* of HealthInsite. These periods are:

- Birth
- Child
- Adolescence
- Adult
- Aged

By clicking on the links, it is possible to obtain automatically the related information resources of these periods. For example, the query generated automatically on Doc CISMef for *Birth* is the following one:

New Born [Keyword] AND Patient [Resource Type] (4)

Today, CISMef-patients is already in use by several patient association libraries. In the near future, we are planning a two-hours training courses on CISMef-patients to enhance its usability.

4. Discussion

Medline-plus [7] developed by the National Library of Medicine (NLM) is a similar catalogue dedicated to the patients. It includes information resources concerning the most current diseases. The Health Topics of Medline-plus have been included after log and queries analyse on Medline. The NLM and National Institutes of Health resources are listed in priority. The number of health topics is larger in Medline-plus than in CISMef-patients (450 vs. 190). The major difference between Medline-plus and CISMef-patients is the structure of the terminology. CISMef-patients and CISMef share the same terminology whereas Medline-plus and Medline do not. Another terminology has been built for Medline-plus. CISMef-patients and CISMef share also the same search tool Doc CISMef. The benefit of sharing the same terminology and search tool, is the possibility to extend the patient query to another one using another resource type than *patient*. For example '*clinical guidelines*' for evidence based medicine resources, or '*education*' for teaching resources. A patient searching information about '*leukaemia*' will have access to '*patient*' resources. She/he may change the Doc' CISMef query into:

leukemia [keyword] AND clinical guidelines [resource type] (5)

leukemia [keyword] AND education material [resource type] (6)

Kogan & al. [6] have identified several problems in patient information retrieval. CISMef-patients, thanks to its useful interface, faces these problems. To formulate queries, the users have only to click on the different links. The articulation of the needs is facilitated. The patients cannot make mistakes or use the wrong medical terms (thanks to the synonyms and to a spelling correction). The resources are written for the patients, so they won't have any problem to understand the information contained in the different resources and in the query results. Finally the content of the catalogue is organised and indexed. We have planned to evaluate the use of CISMef-patients in collaboration with French Patients Associations. It is also necessary to direct the users to information resources of high quality [5]. Using CISMef and CISMef-patients the users don't have to manually filter the information and assess themselves the quality. The CISMef librarians using the Net Scoring criteria [16] are doing this task in the indexing step.

Since 1995 for CISMef, and since 1997 for CISMef-*patients* we use the Dublin Core metadata standard to model the information. Eysenbach[18] points out that Internet metadata standards such as Dublin Core and RDF will allow a machine understandable information. This will improve the intelligent information retrieval and make the Web evolving to a global medical knowledge base and then to the Medical Semantic Web .

5. References

- [1] Eysenbach G, Jadad AR. Consumer Health Informatics in the Internet Age. *Journal of Medical InternetResearch*, 2001; 3(2):e-19.
- [2] Impicciatore P, Pandolfini C, Cosella N, Bonati M. Reliability of Health Information for the Public on the World Wide Web: Systematic Survey of Advice on Managing Fever in Children at Home. *BMJ*, 1997; 314(7098):1875-1877.
- [3] Widman LE, Tong DA. Requests for Medical Advice from Patients and Families to Health Care Providers who Publish on the World Wide Web. *Arch. Intern. Med*, 1997; 157(2):209-212.
- [4] Taylor H, Explosive Growth of 'Cyberchondriacs' Continues.
http://harrisinteractive.com/harris_poll/index.asp?PID=117

- [5] Shepperd S, Charnock D, Gann B. Helping Patients access High Quality Health Information. *BMJ*, 1999; 319:764-766.
- [6] Kogan S, Zeng Q, Ash N, Greenes RA. Problems and Challenges in Patient Information Retrieval. Proc. *AMIA Symp* 2001.
- [7] Miller N, Lacroix EM, Bakus JEB. MEDLINE-plus: Building and Maintaining the National Library of Medicine Consumer Health Web Site. *Bull. Med. Lib. Assoc.*, 2000; 88(1):11-17.
- [8] Darmoni SJ, Leroy JP, Thirion B, Baudic F, Douyère M, Piot J. CISMef a Structured Health Resource Guide. *Methods Inf. Med.*, 2000; 39(1):30-35.
- [9] Darmoni SJ, Thirion B, Platel S, Douyère M, Mourouga P, Leroy JP. CISMef-patients a French Counterpart to MEDLINE-plus. *Journal of the Med. Lib. Assoc.*, 2002; 90(2) 248-253.
- [10] Hudgins J, Agnew G, Brown E. Guetting Mileage Out of Metadata : Applications for the Library. *American Library Association*, 1999.
- [11] Weibel S, Juha H. DC-5 : The Helsinki Metadata Workshop : A report on the Workshop and Subsequent Developments. *D-Lib Magazine*, 1998.
- [12] Darmoni SJ, Thirion B, Leroy JP, Douyère M, Piot J. The use of Dublin Core Metadata in a Structured Health Resource Guide on the Internet. *Bull. Med. Lib. Assoc* ,2001; 83(3): 297-301.
- [13] Eysenbach G. et al. Collaboration for Internet Rating Certification, Labeling and Evaluation of Health Information (MedCIRCLE). *11th WWW Conference*, 2002.
- [14] Medical Subject Headings. National Library of Medicine.
<http://www.nih.gov/mesh/meshhome.html>.
- [15] Koch T. Quality Controlled Subject Gateways: Definitions Typologies, Empirical Overview. *Online Information Review*, 2000; 24(1):24-34.
- [16] Centrale Santé. Net Scoring : Critères de Qualité de l' Information Médicale sur l' Internet.
<http://www.chu-rouen.fr/dsii/puli/critqualV2.html>.
- [17] Darmoni SJ, Thirion B, Leroy JP, Douyere M, Lacoste B, Godard C, Rigolle I, Brisou M, Videau S, Goupy E, Piott J, Quere M, Ouazir S, Abdulrab H. A Search Tool Based on Encapsulated MeSH Thesaurus to Retrieve Quality Resources on the Internet. *Medical Informatics & the Internet in Medicine*, 2001; 26(3):165-178,.
- [18] Eysenbach G. Consumer Health Informatics, *BMJ*, 2000; 320:1713-1716.

Address for correspondence

Lina Fatima Soualmia
 Equipe CISMef DIR CHU de Rouen
 1, rue de Germont
 76031 Rouen Cedex, France.
 e-mail: lina.soualmia@chu-rouen.fr
 url: <http://www.chu-rouen.fr/cismef>