Use of Information and Communication Technologies to retrieve French pre-Residency Examination Program teaching resources on the Internet

**Stefan J. DARMONI [a, b, c], Benoit THIRION [a, c], Benjamin POURCHEZ [a],
Jacques WEBER [c]**

[a] CISMeF, Rouen University Hospital, France
[b] Perception, Information and System Lab, INSA Rouen & Rouen University, France
[c] L@sTICS, Rouen University Hospital & Medical School, France

Abstract

Background: The Internet has become a major source of health information for medical students. Access to accurate information on the Internet is not so easy. Objective: In 2004, a New French Pre-Residency Examination (NFPRE) will be implemented for all the medical students in the 6th year of the curriculum. The goal of this study is to evaluate CISMeF as a tool to provide teaching resources available on the Internet covering NFPRE material. Results: the CISMeF module for the New French Pre-Residency Examination is efficient as it already covers 76% of the program with a precision of 63.5%. Conclusion: This tool should be useful for French medical students.

1. Introduction

Internet tools and services, such as the Web are spreading in every part of the society, including medical schools. It is obvious that they are and will be of the utmost importance to pass on and disseminate knowledge for medical students [1]. Students regularly attend online classes while amphitheatres are nearly empty. Even if the students are now convinced of the interest of Information and Communication Technologies (ICT), most of the French medical teachers are not yet sure that these technologies can be the appropriate means to pass on science and know-how [2]. Nonetheless, medical schools have substantially altered their attitudes with respect to ICT for students [3-4]. It becomes urgent to develop methods and tools that will be appropriate for storing, manipulating and managing teaching material, which will be shareable and easy to (re)use. Moreover, these methods and tools must be based on ICT to maximize the easiness of the handling and the access to these resources to be as large as possible, including local and distant participation.

Access to accurate information on the Internet is not so easy; therefore, there are a great number of directories and search engines available in this new media [5]. However, directories, such as Yahoo [http://www.yahoo.com], or search engines, such as Google [http://www.google.com] do not allow the end-user to obtain a clear and organised range of available useful health information. Therefore, there is a need to develop quality-controlled health subject gateways to disseminate relevant health information. Koch [6] defined quality-controlled subject gateways as Internet services which apply a rich set of quality

measures to support systematic resource discovery. Considerable manual effort is used to process a selection of resources which meet quality criteria and to display a rich description and indexing of these resources with standards-based metadata. Regular checking and updating ensure good collection management. An important goal is to provide a high quality of subject access through resources manually indexed by medical librarians using controlled vocabularies and by offering a deep classification structure for advanced searching and browsing during the information retrieval process.

The objective of CISMeF (French acronym for Catalogue and Index of health resources in French) [7-8] is to describe and index the main health resources in French to assist health professionals and consumers in their search for electronic information available on the Internet. CISMeF is a quality-controlled subject gateway initiated by the Rouen University Hospital (RUH). Its Universal Resource Locator (URL) is http://www.chu-rouen.fr/cismef. CISMeF began in February 1995. In September 2002, more than 10,000 resources have been indexed, with an average of 50 new resources indexed each week. CISMeF has three top priorities in terms of editorial process: teaching resources for medical students (N=1,900), evidence-based medicine information (guidelines and consensus conferences) for healthcare professionals (N=1,000), and consumer health information for patients, their families and more broadly the general public (N=1,400).

2. Objectives

In 2004, a New French Pre-Residency Examination (NFPRE) will be implemented for all French medical students in the 6th year of the curriculum. The material covered by the NFPRE contains 345 questions such as 'question 3: reasoning and decision in medicine. Evidence-based medicine. Therapeutic hazard' and 'question 57: Arthrosis'.

 The goal of this study is to evaluate CISMeF as a tool to provide teaching resources available on the Internet covering NFPRE material. This study is performed under the French Medical Virtual University (FMVU) consortium umbrella (URL: http://www.umvf.org) [10]. This consortium was created to experiment various tools and methods necessary to build a virtual university. Nine medical schools have joined this consortium: Grenoble, Lille, Marseille, Nancy, Nice, Paris V, Paris VI, Rennes and Rouen. FMVU was partially granted by the Health Technologies National Network program of the French Ministry of Research. Eleven working packages were defined, including the 4th working package devoted to automatic and manual indexing. CISMeF is one of the tools tested in this working package. The FMVU consortium is more globally connected with 11 'Digital Campuses' in charge of developing teaching content funded by the French Ministry of Education (each campus developed content for one medical specialty, e.g. medical imaging and microbiology) and a specific e-learning health school also funded by the same Ministry.

3. Material and methods

CISMeF terminology

The CISMeF team is composed of 4 medical librarians, 2 medical informaticians, 1 engineer, 2 PhD students and 1 Master student (all Computer Science students). CISMeF uses two standard tools for organising information: the MeSH (Medical Subject Headings) thesaurus from the US National Library of Medicine (URL: http://www.nlm.nih.gov/mesh/meshhome.html), and several metadata element sets: (a) the Dublin Core metadata format [13] to describe and index all the health resources included in CISMeF, (b) some elements from IEEE1484 Learning Object Metadata for teaching

resources, (c) specific metadata for evidence-based medicine resources which also describe the health content, and (d) the HIDDEL metadata set [9] from the European Union funded MedCIRCLE project (URL: http://www.medcircle.info). HIDDEL will be used to enhance transparency, trust and quality of health information on the Internet.

The MeSH thesaurus contains 20,762 MeSH terms and 83 qualifiers in its year 2002 version as well as nine classification levels. MeSH subheadings allow a focus on a sub-field of a MeSH term, e.g., chloride/toxicity. We also use the French translation of this thesaurus, performed by the French Medlars Centre, the National Institute for Health and Medical Research (INSERM, and more specifically the DISC-DOC Network). We have chosen the MeSH thesaurus because it is the most well-known thesaurus among health professionals, including in France. The CISMeF terminology 'encapsulates' the MeSH thesaurus with two concepts: meta-terms [11] (n=64) and resource types [12] (n=125). A 'meta-term' is generally a medical specialty or a biological science, e.g., cardiology or bacteriology. In most cases, these medical specialities are MeSH terms. The idea of using meta-terms came up to cope with the relatively restrictive nature of MeSH terms. For instance, the queries 'guidelines in cardiology' and 'databases in virology' where cardiology and virology are only MeSH terms get few or no answers. Introducing cardiology and virology as meta-terms is an efficient strategy to get more results. The list of metaterms is available at the following URL: http://www.chu-rouen.fr/ssf/santspeeng.html. The resource types are a generalisation of the publication types of Medline. We have added another concept of 'resource types', taking into account the specific to the resources available on the Internet, such as association, patient information and community networks. The list of resource types is available at the following URL: http://www.chu-rouen.fr/documed/typeeng.html. The keywords, headings and resource types are organised in the form of a hierarchy of terms allowing the most powerful capabilities of this thesaurus: its ability to explode headings to capture narrower terms beneath them in the 'encapsulated' MeSH tree structure. Each meta-term has one semantic link with one or more keywords, subheadings and resource types. Each term can have a set of synonyms, can belong to several trees in the MeSH thesaurus, and have links with several meta-terms. For example the MeSH term 'skin tumour' is associated with the meta-terms 'dermatology' and 'oncology'. A term can be a keyword, a subheading and also a meta-term (eg. Virology). CISMeF uses both MeSH major and minor topics. Indexing an Internet resource in CISMeF with a MeSH major topic means that this MeSH term is very important for that resource (if it is not the case, it will be indexed as a minor topic).

CISMeF and the New French pre-Residency Examination

To help medical student of the FMVU consortium, we developed in July and August 2002 a CISMeF module for the New French pre-Residency Examination (NFRPE). The main objective of this CISMeF module is to automatically map the 1,800 French-speaking teaching resources previously described and indexed in CISMeF to any question of the NFRPE program, optimising the retrieval of quality-controlled teaching resources. These teaching resources are mainly produced by medical schools. The automatic map is based on manual pre-requests performed by the CISMeF chief medical librarian. This task is complex and crucial, that is why it was done by the best CISMeF information scientist. Nonetheless, some requests are rather simple: e.g. for the question 57 'arthrosis', the request is: arthrosis[MeSH term](Major Topic). Most of them are almost impossible to build for medical students: e.g. for question 3 (see above), the request is: (((decision making[MeSH term](Major Topic) or Evidence-Based Medicine [MeSH term](Major Topic) or problem solving[MeSH term](Major Topic) or therapeutic* hazard*

[title/abstract] or ((physician-patient relations[MeSH term](Major Topic) or patient education[MeSH term]) and therapeutics[metaterm]))). Defining requests takes from 10 minutes for the simplest questions to one hour for the most complex questions.

Based on the requests defined by the medical librarian, we developed a generic request for question number (x): ((request of the librarian) or (question (x)) and (education (resource type)). We added a new field called "question(x)" in the CISMeF database in order to indicate that a resource was specifically written for the NFRPE program and more precisely for question (x). Education is a resource type which has several terms below it in the resource type hierarchy (e.g. teaching material, problems based learning, multiple choice quiz, …). For example, for question n° 57 "arthrosis", the generic request is: ((arthrosis[MeSH term](Major Topic) or (question (57)) and (education (resource type)).

Any generic request for one specific question can be limited by the end-user to the resources specifically developed for NFRPE. The limited request is then: (question (x)) and (education (resource type)). Any generic request may also be extended to all the resources indexed in CISMeF covering the question material which are not teaching resources (e.g. clinical guideline, technical report, …). The extended request is then: (request of the librarian) or (question (x)).

Evaluation methodology

Within a classical framework of evaluation, we attempt to find the relevant resources covering NRFPE questions. In computing science, the standard measures of performance are precision & recall, where recall is the ratio between the number of relevant resources extracted by CISMeF and the number of overall relevant resources and the precision is the ratio between the number of relevant resources extracted by CISMeF and the number of overall resources extracted by CISMeF. In information science, the standard measures of performance are silence and noise, where silence is the CISMeF false negative rate and the noise is the CISMeF false positive rate. In this evaluation, we have considered CISMeF as our "gold" standard (reference), thus making it impossible to measure the overall relevant resources. Therefore, we have focused this evaluation study on the precision and the noise. The evaluation was performed on first 88 questions of the program and we have evaluated the number of answers for generic, limited and extended requests. The true positive rate was evaluated by a teacher of the CISMeF team.

4. Results

The Laboratory of Information and Communication Technologies in Health of the Rouen Medical School (French acronym: L@STICS) developed a specific home page for this New French Pre-Residency Examination (NFPRE) with the following main links (URL: http://www.univ-rouen.fr/medecine/pedagogie/internat/): (a) the NFRPE detailed Program of the 345 questions, (b) a correspondence list where the student will find how and when in the Rouen curriculum he will be taught the material of each question, (c) the CISMeF page of clinical cases resources already indexed in the CISMeF database (N=70), (d) (c) the CISMeF page of critical appraisal resources already indexed in the CISMeF database (N=2), and last but not least (e) the CISMeF module for the NFRPE. Points (c) and (d) are important because clinical case and critical appraisal will be the basis of the NFRPE (respectively with 80% and 20% of the final mark). In CISMeF terminology, clinical cases and critical appraisal are both resource types which are located below the 'education' resource type in the hierarchy.

The evaluation of the NRFPE CISMeF module was performed on the 88 pre-requests from

which requests have been generated by the chief medical librarian in September 2002 (see Table 1). The precision of the generic requests is 63.5% (the noise is 36.5%). More important, the coverage of these generic requests which measures the number of questions covered by at least one resource in CISMeF is 85% (76% if taking into account only the true positives). Still very few resources are specifically devoted the NRFPE program (18 resources vs. 375; coverage: 8% vs. 85%). The coverage of the extended requests is 92%. But the average number per question is also too important. Therefore, this extended request should be limited to the questions that cannot be answered with the generic request.

Nonetheless, based on these results, some refinements were added on specific requests because these requests were too noisy: e.g. question 66 with 34 answers and only 12 true positives. Therefore, taking into account the number of CISMeF answers, the refinement mainly consisted in modifying several requests by limiting their scope using more systematically MeSH major topics: the question 66 with a modified request provides now 26 answers.

Table 1: Evaluation of the CISMeF module (88 questions)

	Generic Request (Mean ± standard deviation)	True positives of the generic request (Mean ± standard deviation)	Limited Request	Extended Request
Total of CISMeF teaching resources	375 * (4.26 ± 5.05)	238 * (2.70 ± 2.84)	18	2691 (30.93 ± 47.37)
Number of NRFPE questions with no CISMeF teaching resources	13 (15%)	21 (24%)	81 (92%)	7 (8%)

* precision = 238/375 = 63.5%

5. Discussion

This study measured the CISMeF precision for the new French per residency examination. This measurement should be performed periodically to refine specific requests bearing excessive noise because new teaching resources are periodically included in CISMeF. Nonetheless, in the near future, this specific CISMeF module will be less useful because a lot of teaching resources will specifically be dedicated to this new examination and indexed in CISMeF as such. On the contrary, this CISMeF module will not only map 'old' teaching resources (previously included in the CISMeF database) but also new teaching resources coming from other French speaking countries (e.g. Canada, Belgium and Switzerland) giving the opportunity to medical students to access other resources. Moreover, it was more efficient to develop generic requests for 345 questions instead of indexing 1,800 teaching resources according to these 345 questions. This flexible, less labor intensive approach not only generated good results, but because of its inherent dynamic nature (new resources will automatically be included) is automatically scalable.

We still have a lot of work ahead of us. First of all, the chief medical librarian has to resume the pre-requests for the 345 questions of the NFRPE: this should be achieved in December 2002. We have already planned a formal evaluation of this CISMeF module for the Rouen Medical School students of the fourth and the fifth year of the curriculum in the first semester of 2003. Nonetheless, several students outside the Rouen Medical School may already use it, as the Rouen NFRPE home page appears first as a result of a Google search for 'Residency 2004'.

As a conclusion, the CISMeF module for the new French pre-residency examination is

efficient as it already covers 76% of the program and the precision is 63.5%. This tool should be useful for French medical students.

6. Acknowledgments

7. References

[1] Brunetaud JM, Darmoni SJ, Souf N, Dufresne E, and Beuscart R. A Resource Server For Medical Teaching. *Methods Inf Med* 2001 (in press).

[2] Polyakov A, Palmer E, Devitt PG, and Coventry BJ. Clinicians and computers: friends or foes? *Teach Learn Med* 2000 Spring: 12(2) piii: 91-5.

[3] Ward JP, Gordon J, Field MJ and Lehmann HP. Communication and information technology in medical education. *Lancet* 2001: 357 piii: 792-6.

[4] Chamberland M. Les séances d'apprentissage du raisonnement clinique (ARC). Un exemple d'activité pédagogique contextualisée adaptée aux stages cliniques en Médecine. *Ann. Med. Interne* 1998 : 149 (8) piii: 479-84

[5] Flannery MR. Cataloging Internet resources. *Bull Med Libr Assoc* 1995: 83(2 Suppl) piii: 211-5.

[6] Koch T. Quality-controlled subject gateways: definitions, typologies, empirical overview. *Online Information Review* 2000: 24 (1) piii: 24-34.

[7] Darmoni SJ, Leroy JP, Thirion B, Baudic F, Douyère M and Piot J. CISMeF: a structured Health resource guide. *Meth Inf Med* 2000: 39 (1) piii: 30-5.

[8] Darmoni SJ, Thirion B, Leroy JP, Douyere M, Lacoste B, Godard C, Rigolle I, Brisou M, Videau S, Goupy E, Piott J, Quéré M, Ouazir S, and Abdulrab H A search tool based on 'encapsulated' MeSH thesaurus to retrieve quality health resources on the Internet. *Med Inform Internet Med*. 2001: 26 (3) piii: 65-78.

[9] Eysenbach G, Yihune G, Lampe K, Cross P, and Brickley D. A metadata vocabulary for self- and third-party labeling of health web-sites: Health Information Disclosure, Description and Evaluation Language (HIDDEL). *Proc AMIA Symp* 2001: 169-73.

[10] LeBeux P, Duff F, Fresnel A, Berland Y, Beuscart R, Burgun A, Brunetaud JM, Chatelier G, Darmoni SJ, Duvauferrier R, Fieschi P, Gillois P, Guille F, Kohler F, Pagonis D, Pouliquen B, Soula G, and Weber J. The French Virtual Medical University. *Stud Health Technol Inform* 2000: 77 piii: 554-62.

[11] Thirion B and Darmoni SJ. Simplified access to MeSH Tree Structures on CISMeF. *Bull Med Libr Assoc* 1999: 87 (4) piii: 480-1.

[12] Darmoni SJ, Thirion B. A standard metadata scheme for health resources. *J Am Med Inform Assoc* 2000: 7(1) piii: 108-9.

[13] Darmoni SJ, Thirion B, Leroy JP, Douyère M, and Piot J. The Use of Dublin Core Metadata in a Structured Health Resource Guide on the Internet. *Bull Med Libr Assoc* 2001 July: 89(3) piii: 297-301.

8. Adress for correspondence

Darmoni Stéfan J., Computer and Networks Department, 1 rue de Germont 76031 Rouen Cedex, France
Tel: +33.232.88.88.29; Fax: +33.232.88.88.32
E-mail: Stefan.Darmoni@chu-rouen.fr