



ELSEVIER

journal homepage: [www.intl.elsevierhealth.com/journals/ijmi](http://www.intl.elsevierhealth.com/journals/ijmi)

## Doc'UMVF: Two search tools to provide quality-controlled teaching resources in French to students and teachers

M. Cuggia<sup>a,\*</sup>, S.J. Darmoni<sup>b</sup>, N. Garcelon<sup>a</sup>, L. Soualmia<sup>b</sup>, A. Bourde<sup>a</sup>

<sup>a</sup> EA3888, UFR 140, University of Rennes, France

<sup>b</sup> Laboratory of Medical Informatics, LITIS EA 4051, Institute of Biomedical Research, University of Rouen, France

### ARTICLE INFO

#### Keywords:

Education  
Distance  
Cataloguing  
Quality control

### ABSTRACT

The UMVF project is a federation of medical teaching resources covering 32 medical schools in France. Today, the indexing of these resources is carried out manually by the CISMef team at the University Hospital of Rouen. This indexing is based on MeSH thesaurus. We use a subset of SCORM metadata standard. This choice was defined in collaboration with the French Medical Virtual University consortium (French acronym: UMVF). Currently, with the UMVF searching tool (called Doc'UMVF), medical students can reach more than 3300 resources useful in their curriculum. Doc'UMVF is developed in close collaboration between the medical informatics laboratories of Rennes and Rouen. In this paper we present two complementary searching tools based on different methods and which are integrated and used to improve both the relevance and the coverage rate of the answers. A specific searching module has been built to retrieve specific resources concerning the National Medical Exam ENC ("Examen National Classant") is also available. Nevertheless, due to lack of time, numerous resources are not yet indexed. Therefore we have decided to use also automatic indexing method (Nomindex). This approach will be improved by further research works, resulting from Rouen and Geneva teams. After having built a searching meta-motor, our objective is to develop a meta-tool intended to index the whole set of digital pedagogical resources produced by the UMVF framework. This manual re-indexing will be carried out only for the most important resources (national references), with a more or less fine granularity.

© 2007 Published by Elsevier Ireland Ltd.

## 1. Introduction

Internet tools and services are increasingly present in every aspect of our social and professional activities, including medical studies. As Internet become a major source of health information, it is obvious it will be of the utmost importance to publish and disseminate knowledge for medical students, health professionals, patients and general public [1]. Students regularly attend online classes while face to face lectures are mostly empty when student have clinical hospital work. 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 convinced 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]. Internet is already widely used by residents (91.5%) at least in the USA [5]; in the Rouen Medical School, the 2005 annual survey showed that 82% of the third year medical students have an Internet access at home. It becomes urgent to

\* Corresponding author at: Laboratoire d'Informatique Médicale, Faculté de Médecine de Rennes, Rue du Pr Léon Bernard, 35000 Rennes, France.

E-mail address: [marc.cuggia@univ-rennes.fr](mailto:marc.cuggia@univ-rennes.fr) (M. Cuggia).

1386-5056/\$ – see front matter © 2007 Published by Elsevier Ireland Ltd.

doi:10.1016/j.ijmedinf.2007.01.001

develop methods and tools that will be appropriate for storing, manipulating and managing teaching material, in order to make it shareable and easy to (re) use. Moreover, easy handling and wide access to the resources – both local and distant – are features that must be optimized by using these methods and tools based on ICT.

Access to accurate information on the Internet is not an easy task. There is a great number of search engines, such as Google (<http://www.google.com>) but they do not allow the end-user to obtain a clear and organized range of available useful health information. Therefore many tools and applications have been developed for the healthcare professionals and until recently databases such as Medline were available only to experts. There is a need to develop quality-controlled health subject gateways to display and 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 ensures good quality 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.

In medical information retrieval, there is a need for on line support: when a medical student is searching information, he/she is not always familiar with the medical domain and vocabulary. This implies a bad formulation of a query when he/she is unaware how to express a request, or mistakes when he/she doesn't know the specific medical term to use.

In this context, the objective of Doc'UMVF (<http://www.umvf.org>) is to assist the medical student during the search of electronic teaching resources available on the Internet in French. Doc'UMVF is the official search engine of the French Medical Virtual University (FMVU) Program [7], which is a federation of French medical schools. Its main goal is to share the production and use of teaching resources generated by academic medical teachers. The search may also include virtual Medical Campuses using secure access from several sites to teaching resources written by French medical societies. Doc'UMVF references high quality-controlled information resources written in French. Doc'UMVF is composed of two search tools: one developed in Rennes and one developed in Rouen which are both integrated in one meta engine. We present in this paper some strategies and heuristics to match as much as possible the users' queries with the French adaptation of the MeSH and to optimize precision and recall of the system.

## 2. Methods

### 2.1. Doc'UMVF in Rennes

The aim of this project is to create a meta search engine and to make the searching process easier for students. This engine

must propose an ergonomic search, in order to display first the most relevant retrieved resource. The interface is very simple for students who can search by keywords and medical specialties.

The meta search engine uses several data sources or systems: first the Doc'CISMeF search engine, then UMVF workflow for indexing resources, the accentuated MeSH [7], the Nomindex indexing tool [8] and the Cross-Language Medical Information Retrieval [9] converted in web services.

We designed a differentiated strategy for each search engine: we first look for UMVF and CISMeF and then in PubMed.

#### 2.1.1. For the UMVF and CISMeF course bases

In order to reduce the page loading time, we decided to make a monthly copy in a local database. We load the list of course resources in CISMeF together with their XML description every month for each specialty. Then we reorganize the indexing made for each resource in order to improve the relevance of the results. For each document, we automatically selected the two most important specialties among those proposed by CISMeF. This reindexation is composed of three steps:

- Extract specialties having the maximal weight in CISMeF.
- Among those specialties, select those whose terms are present in the Title of the document and put them as major specialties of the document.
- If no specialty has been selected in step 2 then the ones having the maximum weight are selected as major specialties for the document.

The same operation is done for the document base indexed by the UMVF workflow.

During the search by specialty, the display of results is composed of two parts:

- The most relevant documents: the search specialty is present in the major specialties of the document.
- The less relevant documents: the search specialty is not in the major specialty listed for the document but is present in the minor specialty of the document.

#### 2.1.2. For PubMed

The meta search engine calls the CMLIR-PUBMED web service in order to propose an English translation which is then transmitted for direct search to PubMed. Then the meta-engine collects the XML notices of PubMed and displays them.

#### 2.1.3. The ergonomics for semantic search

We developed the meta search engine to make the searching process easier for students. The interface is very simple for students who can search by keywords and specialties. The engine automatically lists the MESH keywords associated with the natural language query using the NOMINDEX web service.

If the engine does not find any document which corresponds to the query, it will call the automatic misspelling corrector using phonetic proximity (Levenshtein module) and compare it with the local keyword base.

Levenshtein distance (LD) is a measure of the similarity between two strings, which we will refer to as the source string (s) and the target string (t). The distance is the number of deletions, insertions, or substitutions required to transform s into t. For example, if s is “test” and t is “test”, then  $LD(s, t) = 0$ , because no transformations are needed. The strings are already identical. If s is “test” and t is “tent”, then  $LD(s, t) = 1$ , because one substitution (change “s” to “n”) is sufficient to transform s into t.

The greater the Levenshtein distance, the more different the strings are.

Furthermore, every day the meta-engine launches an analysis of the user’s log to analyse the last resource visited by a user before logging off [10,11]. We supposed that this is an interesting resource for the user and therefore we weight it with a high score in a future similar query.

#### 2.1.4. The meta-engine evaluation

We first assessed the capacity of the meta-engine to retrieve indexed documents by CISMef and the ones given by PubMed. The evaluation of the CLMIR with PubMed has been exposed in a previous study [18]. The assessment was based on specialties to measure the relevance of the results, specifically in terms of the ordering of specialties between the meta engine and CISMef.

A survey done in 2002 on search engine user attitudes showed that only 23% of the users went beyond the second page [12]. Another pilot study found that the users looked past the fourth page of results less than 5% of the time [13]. Position ranking in Web-search results, especially on the first few pages, is an important determinant of information accessibility by users. This is why we limited our effort to the 25 first results. For 5 disciplines, we counted the number of documents in terms of noise, silence and relevance in the meta engine and CISMef. The result obtained will enable us to validate the reindexation methods used in the UMVF indexation workflow.

An indexing workflow has been created for the UMVF project. It allows teachers to directly index their course materials with in a web form which automatically gives an accentuated MESH keyword corresponding to the analysis of the description notes of the courses. This is done thanks to the NOMINDEX web service. NOMINDEX is a tool which was developed at the Medical Informatics Laboratory of Rennes (France) and generates MESH keywords from a given sentence or paragraph. Teachers can select, at most, 10 specialties for their courses. Then the workflow produces an XML description using the SCORM standard. This workflow is currently being validated and it will be released for routine use.

## 2.2. Doc’UMVF in Rouen

The search tool developed in Rouen is derived from the CISMef catalogue [14]. Each catalogue resource is indexed by its container using *metadata* and by its contents using the *terms* of the CISMef terminology. Metadata refers to descriptive information about the Web resources and is used to improve information retrieval [15]. In CISMef, we use the Dublin Core (DC) metadata format [16]: each resource is described by eleven of the fifteen elements of DC: author, date,

description, format, language, editor, resource type, rights, subject/keywords and title. For more precision in the resource description, eight elements specific to CISMef were added [17]: institution, city, province, country, target public, access type, cost and sponsorship of the considered resource. The CISMef terminology ‘encapsulates’ the French version of the MeSH thesaurus. However, the MeSH was originally intended to index scientific articles for the Medline database. In order to customise it to the broader field of health Internet resources we have developed several enhancements to the MeSH since 2000. In addition to MeSH keywords and subheadings, the concepts of metaterms (MT) ( $n = 108$ ) and resource types (RT) ( $n = 274$ ) were added. As defined by the DC metadata initiative, a RT is used to categorize the nature of the content of the resource. MeSH (term/subheading) pairs describe the topic of the resource. A MT is generally a medical specialty or a biological science, which has semantic links with one or more MeSH terms, subheadings and RTs (e.g. *cardiology*, *bacteriology*). These medical specialties are in most cases MeSH terms. The idea of the metaterm was established to cope with the relative restrictive nature of these MeSH terms when searching ‘*guidelines in cardiology*’ or ‘*databases in virology*’ where ‘*cardiology*’ and ‘*virology*’ are MTs and ‘*guidelines*’ and ‘*databases*’ are RTs. The RTs are generalisation of the publication types of Medline. The keywords, headings and resource types are organised hierarchically. Compared to the publication types of Medline, the CISMef RTs are more diverse, with specific RTs dedicated to electronic health resources (e.g. *association*, *clinical guidelines*).

CISMef RTs are organised in a hierarchical structure with subsumption relationships. Nonetheless, this list is largely inspired by the MeSH thesaurus as 187 RTs (76%) are deliberately ambiguous because they are also MeSH terms (e.g. *magnetic resonance imaging*). The objective of this ambiguity is to maximise the number of search results (which means the Doc’CISMef search the answers for the MeSH term and the answers for the RT) when the user query contains this kind of ambiguous term. Furthermore, to be as close to a standard as possible, 28 RTs (11%) are also Medline publication types (e.g. *technical report*). Each MT has a semantic link with one or more keywords, headings and RTs. Each term can have a set of synonyms and can belong to several trees.

Many ways of navigation and information retrieval are possible in the catalogue. The most used is the *simple search* (free text interface). It is based on the subsumption relationships. If the query (a word or an expression) can be matched with an existing term of the terminology, then the result of the query is the union of the resources that are indexed by the term, and the resources that are indexed by the terms it subsumes, directly or indirectly, in all the hierarchies it belongs to according to the following formula:  $\bigcup_{i=1}^4 \text{exp}(x)$  (explode function). A query on *Hepatitis* will return as answer all the resources indexed with the descriptor *Hepatitis* but also those indexed with *Hepatitis A*, *Hepatitis B*, etc. If the query cannot be matched, the search is done over the other fields of the metadata and in a worse case a full-text search is carried out. Contrary to Medline, the resource types and the metaterms were voluntary made ambiguous to maximise the recall.

We propose in the following, some enhancements for query matching.

### 2.2.1. Basic natural language processing

The basic natural language processing steps are founded on the following operations.

**2.2.1.1. Query segmentation. Character normalisations.** We apply two types of character normalisation at this step. (1) *Lowercase conversion.* (2) *Deaccenting.*

*Stop words.* We eliminate all the stop words (such as *the, and, when*) in the query. Our stop words list is composed by 1422 elements.

*Exact expression.* We use regular expressions to match the 'exact' expression of each word of the query with the terminology. This step allows to take into consideration the complex terms (composed of more than one word) of the vocabulary and also to avoid some inherent noise generated by the truncations. The query 'accident' is matched with the term 'circulation accident' but not with the terms 'accidents' and 'chute accidentelle'.

*Phonemisation.* The study of the users' queries have shown that a great percent of no answer result from spelling mistakes. We have developed a module that gives the French phonemic conversion of a term. The query is *alzaymer* is replaced by the reserved term *alzheimer* which is orthographically correct. If the phonemic of the query couldn't be matched with the phonemic of a reserved term, a spelling correction is proposed to the user.

*Spelling correction (optional).* The module of spelling correction proposes to the user the reserved term that has a similar phonemic (according to a score and taking into account the possible characters' inversions) with a reserved term. The query is not replaced but only a correction suggestion is proposed to the user. This step is very useful specially in the case of medical vocabulary which contains complex scientific terms.

*Bag of words.* This algorithm search in the user query the greatest set of words that corresponds to a reserved term. The query is segmented. The stop words are eliminated. The other words are phonemised and sorted alphabetically. The different reserved term bags are formed iteratively until possible combinations are exhausted. The query 'therapy of the breast cancer' gives two reserved words: 'therapeutics' and 'breast cancer' (therapy being a synonym of the reserved term therapeutics).

*Heuristics to return documents from the database.* The complex terms (more than 1 word) matching is more requiring than simple terms matching. The CISMeF team policy concerning the queries' rewriting consists in maximizing as much as possible the Doc'CISMeF's recall silence. When all the terms of the query couldn't be recognized as reserved terms, we have implemented five main steps.

*Step 1: the reserved terms.* The process consists of recognizing the user' query expression. If it matches a reserved term of the terminology, the process stops, and the answer of the query is the union of the resources that are indexed by the term, and the resources that are indexed by the terms it subsumes, directly or indirectly, in all the hierarchies it belongs to. If it doesn't match a reserved term, the query is segmented to seek if it contains one or more reserved terms. The query 'enfant asthme' is replaced by (enfant.mr and asthme.mr), where

enfant and asthme are reserved terms (mr). The reserved terms are matched thanks to the *bag of words* algorithm independently of the words' query order. This is useful when the user has not a good knowledge of the terminology specifically at the syntactic level. The expression 'aspergillose allergique bronchopulmonaire' is automatically recognized as being the reserved term 'aspergillose bronchopulmonaire allergique'.

*Step 2: the documents' title.* The search is performed over the other fields of the metadata. The field title of the documents is considered in priority. The stop words are eliminated and the search is realised over the union of the words of the query with a truncation (symbol \*) at the right in the field title (ti).

*Step 3: mixing the reserved terms and the titles.* The system seeks if some words are reserved terms. A new Boolean query is generated with the fields reserved term (mr), if the word is a reserved term, and title (ti) if not. The query 'allergie infantile' is replaced by the Boolean query: (allergie.mr and infantile.ti) because allergie is a reserved term and infantile not.

*Step 4: mixing the reserved terms, all fields and adjacency in the titles.* The search is processed over all the fields (tc) of the documents' metadata for the words that couldn't be recognized as reserved terms union the initial query processed over all the fields with adjacency (at) at  $n$  words with  $n = 5 \times (\text{nb words of the query} - 1)$ . The query 'les problèmes respiratoires des enfants' is replaced by the Boolean query [(enfant.mr and problemes.tc and respiratoires.tc) or (problemes respiratoires enfant.at)]. In this query, the word enfant is recognized as a reserved term because it has the same sonority as the reserved term enfants. The words problèmes and respiratoires are searched over all the fields and the initial query problèmes respiratoires enfants is searched over all the fields with adjacency of 10 which means that these 3 words shouldn't be distant at more than 10 words.

*Step 5: mixing the reserved terms, all fields and adjacency in the plain texts.* A plain text search over the documents with adjacency (ap) of  $n$  words with  $n = 10 \times (\text{nb words of the query} - 1)$  is realised. The query 'bronchite asthmatiforme' is replaced by the Boolean query (bronchite asthmatiforme.ap) where the words bronchite and asthmatiforme shouldn't be distant at more than 10 words in the plain texts of the documents. The plain text search is possible with the Intermedia Text tool of Oracle® 9.i. which required a pre-treatment of the CISMeF corpus (~72h).

To inform the users about their queries' operations, an intuitive scale of interpretation (from Steps 1-5) is available.

### 2.3. Module for French Pre-Residency Examination

Since 2004, a new French Pre-Residency Examination (PRE) is compulsory for all medical students in the sixth year of the curriculum. The goal of this specific module is to provide teaching resources available on the Internet covering PRE material (URL: <http://docCISMeF.chu-rouen.fr/servlets/ECN>). The CISMeF terminology and the PRE CISMeF module are described. To assess the Doc'UMVF PRE module performance in covering PRE program, its precision (number of relevant resources/number of overall resources extracted by CISMeF) and coverage (number of PRE questions covered by at least one resource in the CISMeF gateway) were computed.

Doc'UMVF has defined an interoperability model based on a XML DTD (*Document Type Definition*) to automatically import and export resource records. In the Medical Schools such as Grenoble, Lille and Rennes, teachers supported by webmasters produces resource records. Then records are automatically loaded in Doc'UMVF. Then, the CISMef librarians indexed the resource record using the 'encapsulated' MeSH.

### 3. Results

Here, we present the results of the meta search engine developed in Rennes. These results obtained in terms of pertinence, noise and silence with respect to specialties are really encouraging. The addition of new knowledge bases might be considered and depends on the quality of indexation of the documents. If they are not well indexed, their integration in our meta-engine might require a complete reindexing process. Furthermore, if the data is not well structured, the meta engine will not be able to analyse it properly. It is clear that the quality of the results depends on the quality and adequation of the indexation process. The meta engine can reorganize the documents with a precise objective which is to give the most pertinent document in the first web page (25 referenced URL). In that respect we are grateful to CISMef, which everybody recognizes as a good manual indexing process that we can trust to give objective and well-structured XML metadata.

To evaluate the results, senior physicians put themselves in the role of a medical student, looking for a specific topic or course to review for their exams. Five specialties were selected: cardiology, Foeto-embryology, pneumology, surgery and urology. If a resource can be indexed in say nephrology and urology, we consider it as pertinent for urology. Broadly speaking, at least 60% of the documents were found pertinent in the 25 first returned references as can be seen in Fig. 1.

In Fig. 2 we can see that an average of 77% of the 25 first documents found by the meta engine were relevant.

There-indexing method used for the CISMef documents increases the number of pertinent results found in the first page for a given specialty and is therefore successful.

Though period, we have carried out a comparative study of coverage rate between the both tools using 20 MeSH terms, mostly from the Disease tree. Results are presented in Table 1. The Rouen's tool has a stronger coverage rate (95%) than the Rennes tool (48%).

Thus, these complementary searching tools need to be used in a close cooperation, each one contributing something to the other.

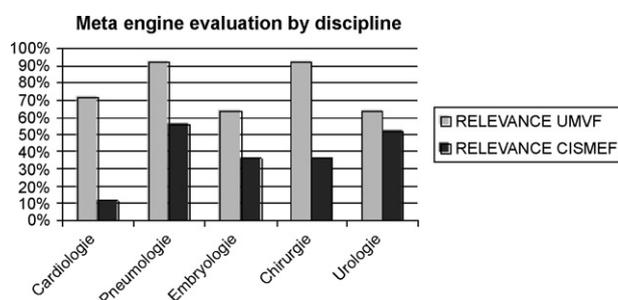


Fig. 1 - Meta engine evaluation by discipline.

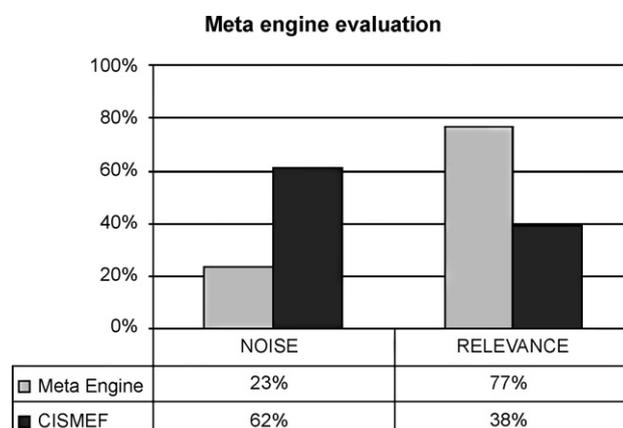


Fig. 2 - Meta engine evaluation.

In March 2006, Doc'UMVF catalogues 3313 teaching resources. Among them, 2696 lectures and 472 problems and exercises (quizzes, problem-based exercises and case reports).

The two Doc'UMVF search tools deeply collaborate. By default, the medical students search the Rennes Doc'UMVF search tool, because it is much simpler to use and because he/she has a better precision (see Table 1). When the Rennes search tool provides no answer or when the end-user wants to extend the query to the Rouen resources using the Rouen algorithm with a better recall, a contextual link is available to the Rouen search tool including the end-user query. When located to the Rouen search tool, a contextual link is available to the Rennes search tool to limit the query to the Rennes resources using the more precise algorithm. When the Rouen search tool provides no answer, a contextual link to Google is available but

Table 1 - Comparative coverage rate

	Doc'UMVF Rennes	Doc'UMVF Rouen	Doc'UMVF Rouen (major)
Postoperative complicator	0	25	0
Chronic disease	12	158	0
Neoplasms	81	326	230
Acute disease	2	182	0
Hypertension	45	76	47
Breast neoplasms	1	25	13
Body weight	0	28	13
Coronary disease	0	34	17
Recurrence	6	31	1
Lung neoplasms	7	14	7
Myocardial infarction	6	33	16
Adenocarcinoma	8	25	4
pain	39	135	97
HIV infections	0	36	10
Asthma	8	26	19
Carcinoma, epidermoid	3	9	4
Liver neoplasm	8	9	6
AIDS	7	26	6
Skin neoplasms	0	12	0
Coverage rate	0.48		0.95

limited as possible to teaching documents, using advanced search and the following query: "inurl:univ OR inurl:ustrasbg OR inurl:ujf-grenoble OR inurl:chups OR inurl:unicaen OR inurl:unilim OR inurl:md.ucl.ac OR inurl:unige.ch OR inurl:ulaval OR inurl:umontreal".

### 3.1. Module for French Pre-Residency Examination

The Doc'UMVF PRE module is efficient as it now covers 100% of the program with a precision of 82.2%. Every working day, an average of 50 end-users (students and teachers) are connected to this module. Our data demonstrates that CISMef is acceptable to guide students' learning and should be a useful teaching resource for the preparation of the French Pre-Residency Examination.

## 4. Discussion and conclusion

A recent Spanish study [19] comparing 6 European health catalogs has shown that CISMef was ranked second after OMNI in terms of precision and recall, mainly because "failure on precision must be due to exhaustive indexing".

In this paper, we have presented strategies to support health information seeking using the CISMef information gateway in the case of free queries that don't match the controlled vocabulary, i.e. that give no answer from the corpus. Simple but essential treatments such as spelling correction are processed online. McCray [20] has also presented strategies for supporting health information seeking. The major difference is in the treatment of the query and specifically in its expansion. We think that this type of query expansion (by relaxing the query) and suggestions to the users may lead them to too many tasks in first, choosing the expanded query and then, in navigating through the documents of the expanded queries to seek the wanted information. The second problem is the exponential growth of the query when it is composed by several words. Another treatment seems to us not necessary at all: the expansion of each word of the query by a set of its synonyms, derivations, inflections. If the query contains a synonym of a reserved term, it should be replaced by the reserved term, which is more precise especially in the context of indexed resources with a controlled vocabulary. The last point concerns the search mode itself which is founded on plain text search vs. indexing terms which is more precise.

In the future, more and more numerous pedagogical resources will be available on the web. Thus, one of our research axes is to develop automated indexing methods. A second challenge is to propose this kind of tool in a transparency way to the scientific community. That is why we have started to implement our tools with web service technologies.

## REFERENCES

[1] J.M. Brunetaud, S.J. Darmoni, N. Souf, E. Dufresne, R. Beuscart, A resource server for medical teaching, *Methods Inform. Med.* 41 (2) (2002) 177-182.

[2] A. Polyakov, E. Palmer, P.G. Devitt, B.J. Coventry, Clinicians and computers: friends or foes? *Teach. Learn. Med.* Spring 12 (2) (2000) 91-95.

[3] J.P. Ward, J. Gordon, M.J. Field, H.P. Lehmann, Communication and information technology in medical education, *Lancet* 357 (2001) 792-796.

[4] M. Chamberland, Les séances d'apprentissage du raisonnement clinique (ARC). Un exemple d'activité pédagogique contextualisée adaptée aux stages cliniques en Médecine, *Ann. Med. Intern.* 149 (8) (1998) 479-484.

[5] J.E. Delzell Jr., R. Weick, M. Weick, How do medical students gather information about residency training programs? *Mo Med.* 100 (2) (2003) 153-154.

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

[7] V. Mary, Indexation de documents bio médicaux et systèmes terminologiques, Thèse, Laboratoire d'Informatique Médicale, Université de Rennes 1, France, 2004.

[8] B. Pouliquen, F. Le Duff, D. Delamarre, M. Cuggia, F. Mougin, P. Le Beux, Managing educational resource in medicine: system design and integration, *Int. J. Med. Inform.* 74 (2-4) (2005) 201-207.

[9] T.D. Tran, N. Garcelon, A. Burgun, P. Le Beux, Experiments in cross-language medical information retrieval using a mixing translation module, *Medinfo 2004* (2004) 946-952.

[10] D. Chen, H.F. Orthner, S.M. Sell, Personalized online information search and visualization, *BMC Med. Inform. Decis. Making* 5 (2005) 6.

[11] J. Crowell, Q. Zeng, L. Ngo, E.M. Lacroix, A frequency-based technique to improve the spelling suggestion rank in medical queries, *J. Am. Med. Inform. Assoc.* 11 (2004) 179-185.

[12] R. Greenspan, Search engine usage ranks high. URL: <http://cyberatlas.internet.com/markets/advertising/article/0,5941,1500821,00.html> (accessed Jun 16, 2003).

[13] C.R. Richardson, P.J. Resnick, D.L. Hansen, H.A. Derry, V.J. Rideout, Does pornography-blocking software block access to health information on the Internet? *JAMA* 288 (22) (2002) 2887-2894.

[14] S.J. Darmoni, J.P. Leroy, B. Thirion, F. Baudic, M. Douyère, J. Piot, CISMef: a structured health resource guide, *Methods Inform. Med.* 39 (1) (2000) 30-35.

[15] J. Hudgins, G. Agnew, E. Brown, Guetting Mileage Out of Metadata: Applications for the Library, American Library Association, 1999.

[16] M. Dekkers, S. Weibel, State of the Dublin core metadata initiative, *D-Lib. Mag.* V9 (2003) N40.

[17] B. Thirion, G. Loosli, M. Douyère, S.J. Darmoni, Metadata element set in a quality-controlled subject gateway: a step to an health semantic web, *Stud. Health Technol. Inform.* 95 (2003) 707-712.

[18] M. Douyère, L.F. Soualmia, A. Névéol, et al., Enhancing the MeSH thesaurus to retrieve French online health resources in a quality-controlled gateway, *Health Inform. Lib. J.* 21 (4) (2004) 253-261.

[19] F. Abad Garcia, A. Gonzalez Teruel, P. Bayo Calduch, R. de Ramon Frias, L. Castillo Blasco, A comparative study of six European databases of medically oriented Web resources, *J. Med. Lib. Assoc.* 93 (4) (2005) 467-479.

[20] A.T. McCray, N.C. Ide, R.R. Loane, T. Tse, Strategies for supporting consumer health information seeking, *Medinfo* (2004) 1152-1156.