

Enhancing the MeSH thesaurus to retrieve French online health resources in a quality-controlled gateway

Magaly Douyère*, Lina F. Soualmia*†, Aurélie Névéal*†, Alexandrina Rogozant, Badisse Dahamna*, Jean-Philippe Leroy*, Benoît Thirion* & Stefan J. Darmoni*†, *CISMeF, Rouen University Hospital, †Perception and Information Systems Laboratory CNRS, Medical School, University of Rouen, Rouen, France

Abstract

The amount of health information available on the Internet is considerable. In this context, several health gateways have been developed. Among them, CISMeF (Catalogue and Index of Health Resources in French) was designed to catalogue and index health resources in French. The goal of this article is to describe the various enhancements to the MeSH thesaurus developed by the CISMeF team to adapt this terminology to the broader field of health Internet resources instead of scientific articles for the MEDLINE bibliographic database. CISMeF uses two standard tools for organizing information: the MeSH thesaurus and several metadata element sets, in particular the Dublin Core metadata format. The heterogeneity of Internet health resources led the CISMeF team to enhance the MeSH thesaurus with the introduction of two new concepts, respectively, resource types and metaterms. CISMeF resource types are a generalization of the publication types of MEDLINE. A resource type describes the nature of the resource and MeSH keyword/qualifier pairs describe the subject of the resource. A metaterm is generally a medical specialty or a biological science, which has semantic links with one or more MeSH keywords, qualifiers and resource types. The CISMeF terminology is exploited for several tasks: resource indexing performed manually, resource categorization performed automatically, visualization and navigation through the concept hierarchies and information retrieval using the Doc'CISMeF search engine. The CISMeF health gateway uses several MeSH thesaurus enhancements to optimize information retrieval, hierarchy navigation and automatic indexing.

Introduction

The amount of health information available on the Internet is considerable and is growing quickly: Eysenbach & Kohler¹ have recently estimated that 12 million health-related searches are being conducted on the Web every day, which is roughly the same number of searches that have been conducted on the NLM Medlars system in 1996 in a full year.

Information retrieval remains problematic: users are now experiencing huge difficulties in finding precisely what they are looking for among the billions of documents available online, in spite of existing tools. Generic search engines (e.g. Google <http://www.google.com>) or generic catalogues (e.g. Yahoo <http://www.yahoo.com>) cannot solve this problem efficiently because they usually offer a selection of documents that turns out to be either too large or ill-suited to the query.

In this context, several health gateways have been developed. We propose a hierarchy of these gateways in three grades:

Correspondence: Stefan J. Darmoni, CISMeF, Rouen University Hospital, 1 rue de Germont, 76031 Rouen Cedex, France. E-mail: stefan.darmoni@chu-rouen.fr

- grade 1: search engine, generalist or more specialized ones, such as MedHunt (<http://www.hon.ch/>);
- grade 2: catalogue and index without thesaurus, such as MedWebPlus-Us (<http://www.medwebplus.com/>) and HealthWeb-Us (<http://healthweb.org/>);
- grade 3: catalogue and index with thesaurus, such as the UMLS (Unified Medical Language System) metathesaurus² and the MeSH thesaurus.³ The latter thesaurus is used in the following Health catalogues: DDRT (Diseases, Disorders and Related Topics (<http://www.mic.ki.se/Diseases/index.html>), Medical Library and Medical Information Center, Karolinska Institute, Stockholm, Sweden), CliniWeb⁴ (<http://www.ohsu.edu/clinweb/>), Oregon Health Sciences University-Us, OMNI (Organizing Medical Networked Information-Uk) (<http://omni.ac.uk/>)⁵ HON (Health on the Net-Ch)⁶ Healthinsite-Au (URL: <http://www.healthinsite.gov.au/>)⁷ and CISMef (Catalogue and Index of Health Resources in French-Fr) (<http://www.chu-rouen.fr> or <http://www.cismef.org>).⁸

Koch⁹ defined quality-controlled subject gateways as Internet services that apply a comprehensive 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 an extensive description and indexing of these resources with standards-based metadata. Regular checking and updating ensure optimal collection management. The main goal is to provide a high quality of subject access through indexing resources using controlled vocabularies and by offering a deep classification structure for advanced searching and browsing.

Objective

The goal of this article is to describe the various enhancements to the MeSH thesaurus that the CISMef team has been developing in order to adapt this terminology to the broader field of health Internet resources instead of scientific articles for the MEDLINE bibliographic database which the MeSH thesaurus was originally built for. We will describe the various properties of the 'enhanced' MeSH thesaurus in information retrieval, categorization and automatic indexing.

The CISMef project

The CISMef project was initiated in February 1995. As opposed to Yahoo, CISMef is a quality-controlled health gateway cataloguing the most important and quality-controlled sources of institutional health information in French in order to allow end-users to search them quickly and precisely. It is manually maintained. CISMef indexes a great variety of resources ($n \approx 13\,000$) but has three main topics: guidelines for health professionals ($n \approx 1500$), teaching material for students ($n \approx 3000$), and consumer health information ($n \approx 1900$). A resource can be a website, web pages, documents, reports and teaching material: any support that may contain health information.

Each of the following steps proposed by Koch, which characterize a typical quality-controlled subject gateway, are implemented in CISMef:

- selection and collection development, based on the Net Scoring, a list of 49 criteria to assess the quality of health information (<http://www.chu-rouen.fr/netscoring>) and more recently on the EU-funded MedCIRCLE project (<http://www.medcircle.info>);¹⁰
- collection management;
- intellectual creation of metadata (carried out by experts);
- resource description (an extensive and documented metadata set);
- resource indexing (using a controlled vocabulary system).

Enhanced MeSH terminology

CISMef uses two standard tools for organizing information: the MeSH (medical subject heading) thesaurus from the US National Library of Medicine and its French translation by the French Medlars Center (French National Institute of Health), and several metadata element sets,¹¹ in particular the Dublin Core metadata format.¹² This structure enables us to place the project at an overlap between the actual informal Web and the forthcoming Semantic Web.¹³

The Semantic Web is an infrastructure that has to be built. It aims at creating a Web where information semantics are represented in a form that can be understood by human as well as machines,

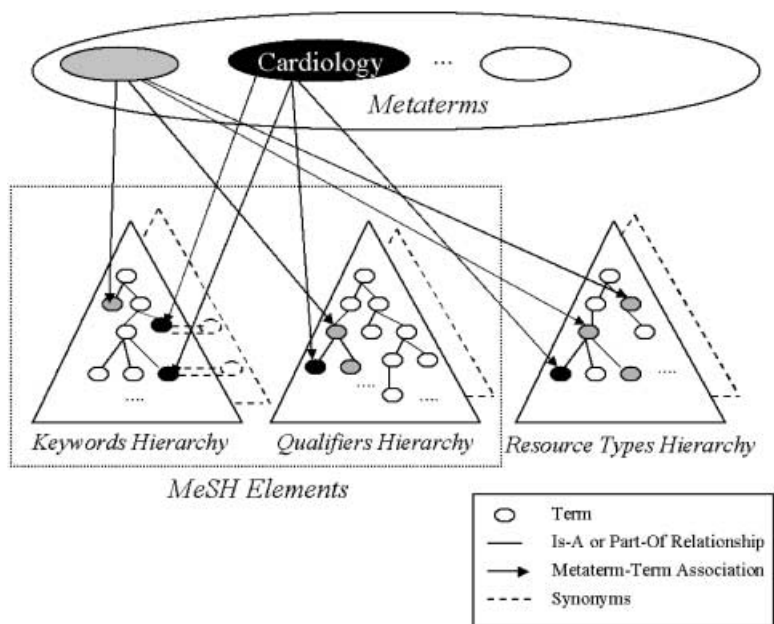


Figure 1 The CISMeF terminology structure

better enabling computers and people to work in co-operation. One of its advantages is to bring sufficient information on the resources, by adding annotations in the form of *metadata* and to describe formally and significantly their content according to an ontology. This infrastructure must be formalized. The current Web is informal: it is mainly composed of html pages hand-written or generated automatically for only human treatment. Ontologies and metadata are two major components for the construction of the Semantic Web.

The CISMeF terminology constitutes the main component of the catalogue and the information retrieval process. The MeSH was selected because it meets the aims of medical librarians and it is well known by the health professionals. Approximately 22 000 keywords (e.g. *abdomen, hepatitis*) and 84 qualifiers or subheadings (e.g. *diagnosis, complications*) compose the MeSH thesaurus in its 2003 version.¹⁴ These concepts are organized into hierarchies going from the most general at the top of the hierarchy to the most specific at the bottom. For example, the keyword *hepatitis* is more general than the keyword *hepatitis viral A*. The qualifiers, also organized into hierarchies, allow specification of which particular aspect of a keyword is addressed. For example, the association of the keyword *hepatitis* with the qualifier *diagnosis* (noted *hepatitis/diagnosis*) restrict the *hepatitis* to its *diagnosis* aspect.

The 'is-a' relations between concepts are extracted from the MeSH text files to define the subsumption relationships in the CISMeF keywords hierarchy ($n \approx 9400$; $\approx 42.7\%$ of the MeSH thesaurus).

The heterogeneity of Internet health resources and the great specificity of MeSH keywords (which make it difficult to refer broadly to a medical specialty), led the CISMeF team to enhance the MeSH thesaurus with the introduction of two new concepts, respectively, resource types and metaterms (see Fig. 1).

CISMeF resource types are a generalization of the publication types of MEDLINE. Types which are specific of the health resources available on the Internet have been added, such as association, patient information, community networks. The controlled list of resource types ($n = 145$) is available at <http://www.chu-rouen.fr/documed/typeeng.html>. As with keywords, qualifiers and publication types, resource types are organized into hierarchies. A resource type describes the nature of the resource and MeSH keyword/qualifier pairs describe the subject of the resource. For example, in the case of a clinical guideline about carbon monoxide intoxication, 'carbon monoxide poisoning' is the MeSH keyword and 'clinical guidelines' is the resource type.

A metaterm is generally a medical specialty or a biological science (e.g. cardiology or bacteriology) selected by the CISMeF chief librarian. For each

metaterm ($n = 69$), one semantic link was created with one or more MeSH keywords, qualifiers and resource types. For example, the metaterm *psychiatry* is associated with the MeSH keywords *psychiatry* and *psychiatric hospital* that belong to a completely different tree structure within the MeSH and also with the CISMef resource type *mental health dispensary*.

In fact, metaterms have been created to optimize information retrieval in CISMef and to overcome the relatively restrictive nature of MeSH keywords. For instance, the queries 'guidelines in cardiology' and 'databases in psychiatry' where *cardiology* and *psychiatry* are the only MeSH keywords get few or no answers. Introducing *cardiology* and *psychiatry* as meta-terms is an efficient strategy to get more results because instead of exploding one single MeSH tree (e.g. *psychiatry* as a MeSH keyword), using metaterms results in an automatic expansion of the queries by exploding other related MeSH or CISMef trees as well as the current tree. (e.g. *psychiatric hospital* as a MeSH keyword or *mental health dispensary* as a resource type will be exploded in the case of the *psychiatry* query). The list of metaterms is available at <http://www.chu-rouen.fr/ssf/santspeeng.html>.

Thanks to the VUMef consortium¹⁵ in charge of extending the French in the UMLS metathesaurus, the CISMef team has added over 1000 synonyms, specially acronyms, several 'see also' relations between MeSH keywords and already translated the definitions of 10% of the MeSH thesaurus. To be freely available to the academics, this work needs to be validated by the French Medlars Center (French Institute of Health).

Terminology exploitation

Resource indexing and categorization. The CISMef terminology (metaterms, keywords, qualifiers and resource types) is exploited for several tasks: resource indexing performed manually, resource categorization performed automatically, visualization and navigation through the concept hierarchies and information retrieval using the Doc'CISMef search engine. Each catalogue resource is indexed using the vocabulary of the terminology.

Categorization is designed to enhance resource description by organizing content description so

as to enable the reader to grasp quickly and easily what a resource is about and what are the main topics discussed in it. Resource categorization is performed automatically. Practically, in CISMef, this type of categorization lists the medical specialties relevant to a resource by decreasing order of their importance in the text. Using heuristics and a categorization algorithm¹⁶ the related specialties to a resource are deduced thanks to the existing semantic links between (metaterm, keyword) (metaterm, qualifier) and (metaterm, resource type) and are ranked according to their level of importance (major/minor topic). An evaluation of this algorithm on a random set of 123 resources described by Neveol *et al.*⁶ gave very satisfying results (81% precision and 93% recall)¹⁶ and led us to use the algorithm as an automatic classification method for CISMef resources.

Information retrieval and navigation. The navigation through the terminology, thanks to alphabetical and thematic indices, allows the user to know the terms that represent the concepts used in the domain and also their positions in the different hierarchies. Each term has its own web page, and a set of links enables the user to retrieve, by preformatted queries, all the resources that are related to this term. He can also restrict the search according to his/her profile: resources intended for health professionals, for students, for patients and for the general public, respectively, with the following resources types: 'guidelines', 'education', and 'patient'. The CISMef team also developed a MeSH navigation (in French) very similar to the MeSH navigation (in English) developed by the PubMed website (<http://pubmed.gov>). This MeSH navigation is available from any MeSH term page in the CISMef website. Then, the end-user may learn the various subsumption relationships existing between MeSH terms, e.g. '*abetalipoproteinemia*' belongs to the following hierarchies: '*hemic and lymphatic diseases*', '*congenital, hereditary, and neonatal diseases and abnormalities*', '*nutritional and metabolic diseases*', and '*nervous system diseases*'. Medical students from the Rouen Medical School are delighted to find such subsumption relationships among MeSH terms ... in French.

The other and main utility of the terminology is its exploitation by the Doc'CISMef search

engine.¹⁷ Different search modes are available. ‘Simple search’ is done via an interface in which the user can type queries in natural language (French or English, with or without accents, in capital letters or not). ‘Advanced search’ is a more precise search: it uses frames and drop-down lists and different attributes from the CISMef metadata set (keywords, titles, year) and can be combined with Boolean operators (AND, OR, EXCEPT). CISMef training sessions emphasize how to use the Advanced search to perform a robust query. ‘Logical search’ is done with Boolean operators and a specific query language with particular characters. The Logical search is mainly intended for medical librarians.

Currently the *simple search* is based on subsumption relationships. If the query (a word or an expression) can be matched with an existing concept of the CISMef terminology which ‘encapsulates’ the MeSH thesaurus (metaterms, keyword, qualifier, resource type), then the result of the query is the union of the resources that are instances of the concept, and the resources that are instances of the concept it subsumes, directly or indirectly, in all the hierarchies it belongs to, as shown in the following equation:

$$\bigcup_{i=1}^4 \text{exp}(x)$$

where i is the level of the CISMef terminology, x is the concept and exp is the explode function. For example, if the end-user enters the term ‘virology’ in English or ‘virologie’ in French, the search will be performed on the exploded meta-term ‘virology’, on the exploded MeSH term ‘virology’ and on the exploded qualifier ‘virology’. If the query cannot be matched to a concept of the terminology, the search is done over the other fields of the CISMef metadata set.

More recently, the CISMef team has introduced a more complex algorithm to optimize information retrieval specifically to reduce the silence (see Fig. 2). The principle of the algorithm is the following: given a query Q (composed of n words $q_1 \dots q_n$) that returns no answer because it doesn’t match any term of the terminology, our objective is to find the best match with the terminology. To do that, the query Q is segmented into $q_1 \dots q_n$ words. The stop words (e.g. *the, before, with* ...) are

eliminated and a transformed query Q' is obtained. For each word q_i , the algorithm tests whether q_i belongs to the terminology. If so, the final Boolean query is: q_i (reserved word). If not, the final Boolean query is: q_i (all fields). An ‘all fields’ search is performed over all the fields of the metadata (including *title, abstract, ... etc.*). All these modifications of the original query Q are performed automatically without the intervention of the user.

Doc’CISMef is interoperable with the MEDLINE bibliographic database via the PubMed website: the CISMef query is automatically transformed in the PubMed syntax. In the worst case, a full-text search is carried out in the CISMef corpus. But, as said before, it is not an optimized solution and this kind of search requires a good knowledge of the medical domain, which is not obvious for any user. This algorithm is under evaluation to measure whether the returned answers, if there are some answers, correspond to the initial query without noise.

Another way to enhance information retrieval within the catalogue is to perform an interactive query expansion with the user (see Fig. 3). All the terms of the terminology that contain the query are proposed to the user and the related keywords that are related with the *see also* relationships. For example if the query is ‘diabetes’ the returned propositions that contain the query are: ‘diabetes, autoimmune’, ‘diabetes mellitus, adult-onset’, ‘diabetes mellitus, brittle’, ‘diabetes mellitus, gestational’, ‘diabetes mellitus, ketosis-prone’, ‘diabetes mellitus, type I’ ... etc. The other propositions correspond to the *see also* related terms of the terminology. In our example of ‘diabetes’ these terms are: ‘blood glucose self-monitoring’, ‘hypoglycemic agents’, ‘diabetic diet’ ... etc. All these propositions could help the user to expand its query and also to have a global view of its corresponding ‘neighbours’.

Automatic indexing. As of December 2003, the indexing of new resources in the CISMef catalogue is still performed manually (55 new resources per week). In order to reduce the manual indexing processing time for the CISMef medical librarians, a PhD candidate (AN) is working on the development of an automatic indexing system that should meet the current manual indexing standards. In fact, the system will be customized

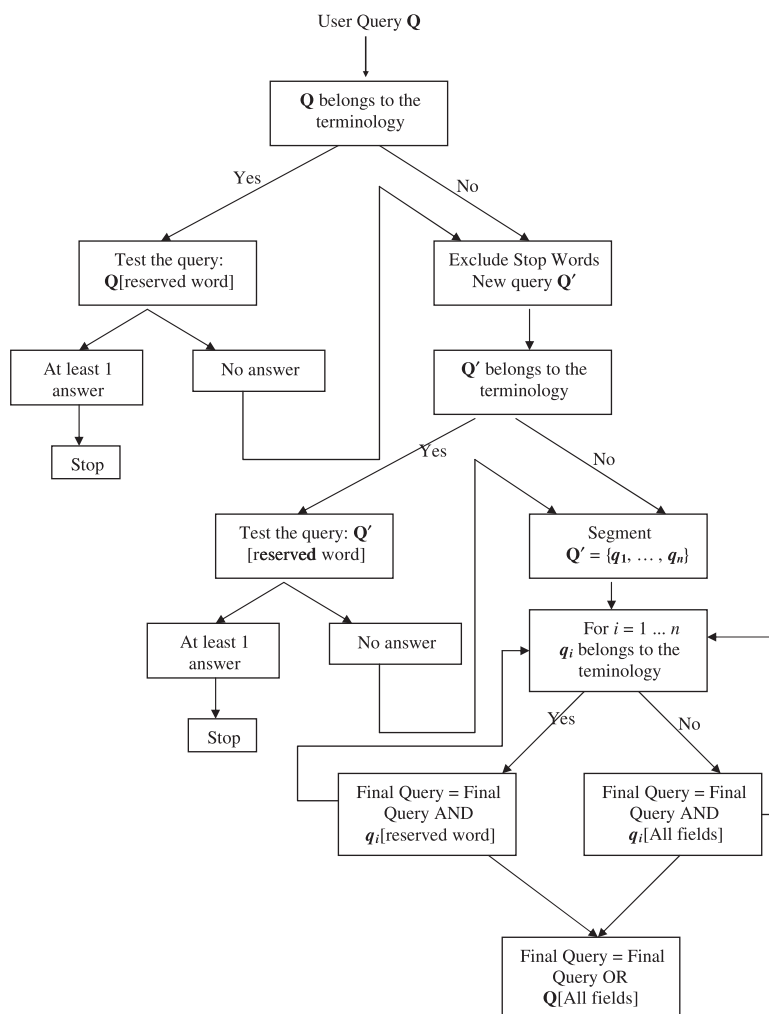


Figure 2 The Doc'CISMeF simple search process

to include the following features: MeSH keywords and keyword/qualifier *pairs* are to be retrieved, as opposed to single MeSH terms as it is the case for the existing automatic indexing systems. Then, each keyword (or keyword/qualifier pairs) is allotted a 'minor' or 'major' weight, according to the importance of the concept it refers to in the resource. The issue of check tag selection will also be addressed.

More specifically, the terminology information at our disposal can be used in various ways. For instance, if several MeSH terms belonging to the same MeSH hierarchy are possible indexing candidates, only the more precise one(s) should be selected, and it (they) should also gain from the weight of the broader term(s). Thus, if the three terms hepatitis, hepatitis A, and hepatitis B occur in a given resource, the hierarchical relationships

between hepatitis and hepatitis A on the one hand and hepatitis and hepatitis B on the other will result in attributing augmented scores to hepatitis A and hepatitis B which are the more precise terms, whereas hepatitis will not appear on the candidate keyword list anymore. The CISMeF terminology also plays an important part in extracting relevant textual elements (i.e. terms or expressions) from the resource to be mapped to MeSH indexing terms. Such textual elements include the MeSH terms (e.g. 'sujetâgé'—aged) and their inflected forms (e.g. 'sujetsâgés') as well as MeSH and CISMeF synonyms of these terms (e.g. 'personneâgée'—elderly), and specific indexing strategies defined by the chief librarian (e.g. the expression 'vaccination against Disease D' should be mapped to the MeSH pair Disease D/ prevention and control). Finally, it is also important

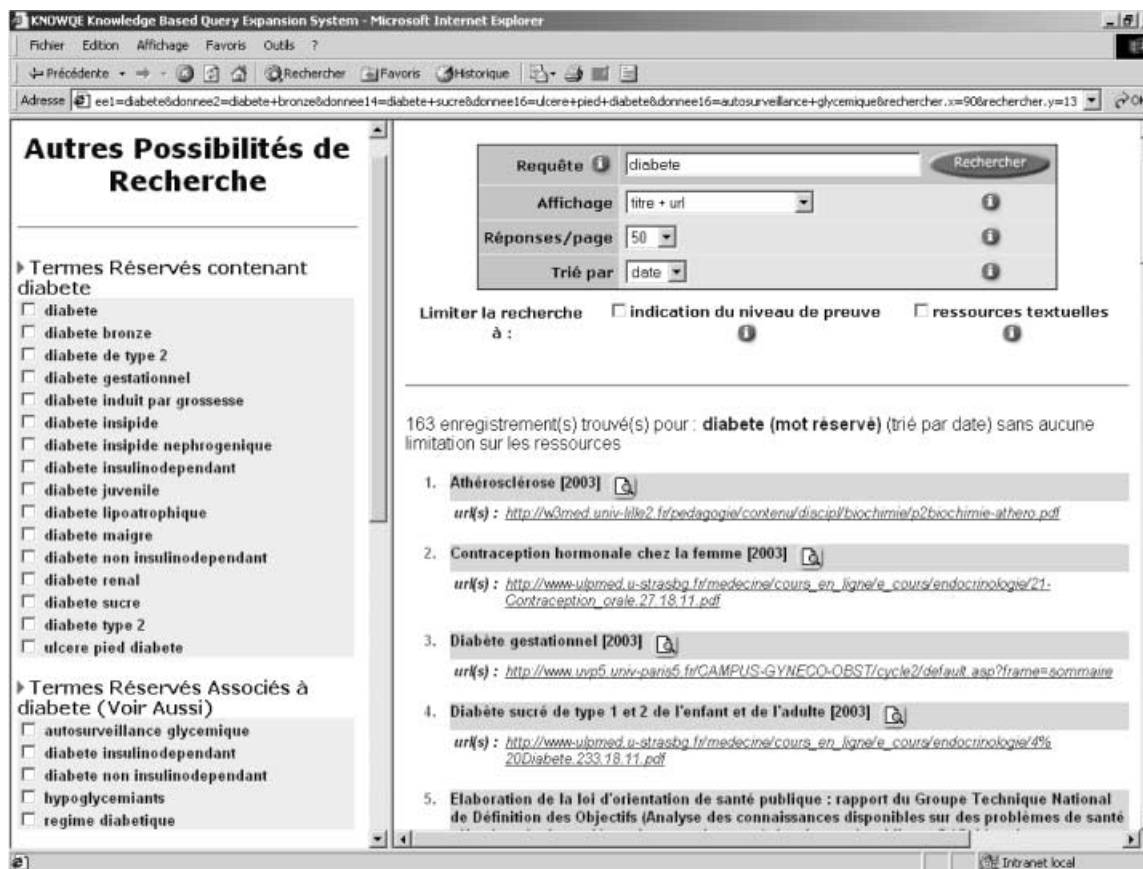


Figure 3 The interactive query expansion prototype

to note that the resources indexed in CISMef are described with a number of MeSH keywords (or pairs) ranging from 0 (e.g. hospital websites) to a few dozens. Hence, a method adapting the length of the indexing list to each resource needs to be developed.

In order to provide a comprehensive description and retrieval of the resources, within the CISMef catalogue, it is necessary to include images both in the indexing and retrieval procedures. Hence, a PhD candidate will start his thesis on image indexing and coupled text-image retrieval in January 2004.

Other features: major topics and search strategies. Any search could be limited to Major Topics of each level of the CISMef terminology. Major Topics exist in the MEDLINE database and the CISMef catalogue for keywords and qualifiers. A term is said to be 'major' if the concept it represents is discussed throughout the whole document or, on the contrary, 'minor' if it is

referred to only in a few paragraphs. Major terms are marked in MEDLINE & CISMef by a star.

In CISMef Major Topics are extended to resource types and metaterms. This task is manually performed by CISMef medical librarians for resource types, and automatically performed for metaterms: a metaterm is 'major' for a CISMef resource if, and only if, at least one keyword, qualifier or resource type semantically linked to this metaterm is major for the same CISMef resource (otherwise, the metaterm is minor).

An other kind of pre-formatted queries has been modelled: the *search strategies*. A search strategy is a medical concept defined by a Boolean expression composed of several other concepts of the terminology. For example: we defined the following search strategies for 'dental surgery' and 'urological surgery':

dental surgery = dentistry, operative ((keyword) OR tooth extraction (keyword)) OR (tooth replantation

(keyword) OR ((dentition(keyword) OR tooth diseases(keyword)) AND surgery(qualifier)); AND urological surgery = urology (metaterm) AND (surgical procedures, operative (keyword) OR surgery (qualifier)).

Discussion and conclusion

We have shown in this article the various properties of the MeSH thesaurus enhancements developed by the CISMef team to optimize information retrieval, hierarchy navigation and automatic indexing.

There are a few drawbacks to these increasingly specific enhancements MeSH thesaurus. As a 'simple search' is being processed, if no answers are produced with the usual method (see Fig. 2), the original query is automatically modified, no notification having been sent to the user.

Chances are that the user will be at a loss, as he won't be able to know which query was actually processed. The results appearing are not necessarily produced by the original query. We are fully aware of the situation and, to avoid confusion, we intend to present the results so as to show the user:

- that his query didn't produce any answers;
- that a modified query can produce answers;
- the answers produced by said query (which may be noisy).

However, the query modifications thus engineered may lead to high noise. Besides the evaluation of the categorization algorithm reported on in¹⁶ information retrieval within the CISMef catalogue using MeSH information only will be formally compared in terms of precision and recall to the information retrieval using CISMef enhancements. The first evaluation will aim at measuring the benefit of the metaterm concept.

The MeSH thesaurus is largely used by the most important catalogues in health.³⁻⁷ The use of the UMLS metathesaurus should improve significantly these three different functionalities (information retrieval, hierarchy navigation and automatic indexing) as the number of concepts is 1.2 million in the UMLS metathesaurus versus 22 000 in the MeSH thesaurus. In order to increase the search precision, we need to add semantic links between concepts to extend the hierarchical links. Nonetheless, such improvements will have to wait until the deliver-

ables of the VUMef consortium¹⁵ in charge of extending the French in UMLS are released.

We have already several perspectives tracks, in particular the optimization of the Doc'CISMef search tool to allow 'intelligent information retrieval'. A PhD candidate (LS) developed in particular the following KnowQuE (Knowledge-based Query Expansion) prototype system¹⁸ which includes:

- a morphological knowledge base in cooperation with Zweigenbaum and Grabar, which will benefit from the UMLF consortium in charge of developing the French Specialist Lexicon,¹⁹ e.g. the query *asthmatic children* will be derived into *asthma AND child*;
- a knowledge base of association rules extracted using the data mining knowledge discovery process, e.g. breast cancer/diagnostic \Rightarrow mammography or hepatitis/prevention and control \Rightarrow hepatitis vaccines;
- a formalized CISMef terminology using the OWL²⁰ language to benefit of the advantages of its powerful reasoning mechanisms.

References

- 1 Eysenbach, G. & Köhler, C. What is the prevalence of health-related searches on the world wide web? Qualitative and quantitative analysis of search engine queries on the Internet. *Proceedings of the AMIA Annual Fall Symposium* 2003, 225-9.
- 2 Humphreys, B. L., Lindberg, D. A., Schoolman, H. M. & Barnett, G. O. The Unified Medical Language System: an informatics research collaboration. *Journal of the American Medical Informatics Association* 1998, 5, 1-11.
- 3 Nelson, S. J., Johnson, W. D. & Humphreys, B. L. Relationships in medical subject headings. In: Bean & Green (eds). *Relationships in the Organization of Knowledge*. Dordrecht: Kluwer Academic Publishers, 2001, 171-84.
- 4 Hersh, W. R., Brown, K. E., Donohoe, L. C., Campbell, E. M. & Horacek, A. E. CliniWeb: managing clinical information on the world wide web. *Journal of the American Medical Informatics Association* 1996, 3, 273-80.
- 5 Norman, F. Organising medical networks' information (OMNI). *Medical Informatics* 1998, 23, 43-51.
- 6 Boyer, C., Baujard, O., Baujard, V., Aurel, S., Selby, M. & Appel, R. D. Health on the net automated database of health and medical information. *International Journal of Medical Informatics* 1997, 47, 27-9.
- 7 Deacon, P., Smith, J. B. & Tow, S. Using metadata to create navigation paths in the HealthInsite Internet gateway. *Health Information and Libraries Journal* 2001, 18, 20-9.

- 8 Darmoni, S. J., Leroy, J. P., Baudic, F., Douyère, M., Piot, J. & Thirion, B. CISMef: a structured health resource guide. *Methods of Information in Medicine* 2000, **39**, 30–5.
- 9 Koch, T. Quality-controlled subject gateways: definitions, typologies, empirical overview. In: *Subject Gateways*. Special issue. *Online Information Review* 2000, **24**, 24–34.
- 10 Mayer, M. A., Darmoni, S. J., Fiene, M., Eysenbach, G., Köhler, C. & Roth-Berghofer, T. MedCIRCLE—modeling a collaboration for Internet rating, certification, labeling and evaluation of health information on the semantic world wide web. In: Surjan, G., Engelbrecht, R. & McNair, P. (eds). *Proceedings of MIE 2003, Eighteenth International Congress of the European Federation for Medical Informatics*. Amsterdam: IOS Press Publisher, Stud Health Technol Inform, 2003, **95**: 667–72.
- 11 Thirion, B., Douyère, M., Loosli, G. & Darmoni, S. J. Metadata element set in a quality-controlled subject gateway: a step to a health semantic web. In: Surjan, G., Engelbrecht, R., McNair, P. (eds). *Proceedings of MIE 2003, Eighteenth International Congress of the European Federation for Medical Informatics*. Amsterdam: IOS Press Publisher, Stud Health Technol Inform, 2003, **95**, 707–12.
- 12 Weibel, S. L. & Koch, T. 2000. The Dublin core metadata initiative. *D-Lib Magazine*. Available from: <<http://www.dlib.org/dlib/december00/weibel/12weibel.html>>.
- 13 Berners-Lee, T., Heudler, J. & Lassila, O. The semantic web. *Scientific American* 2001, **284**, 34–43.
- 14 Medical Subject Headings. Home page. 2003. Available from: <<http://www.nlm.nih.gov/mesh/meshhome.html>> (accessed 8 March 2004).
- 15 Darmoni, S. J., Jarrousse, E., Zweigenbaum, P., Le Beux, P., Namer, F., Baud, R., Joubert, M., Vallée, H., Côté, R. A., Buemi, A., Bourigault, D., Recourcé, G., Jeanneau, S. & Rodrigues, J. M. VUMeF: extending the French involvement in the UMLS metathesaurus. *AMIA* 2003, 824.
- 16 Neveol, A., Soualmia, L. F., Douyère, M., Rogozan, A., Thirion, B. & Darmoni, S. J. Using CISMef MeSH ‘encapsulated’ terminology and a categorization algorithm for health resources. *International Journal of Medical Informatics* 2004, **73**, 57–64.
- 17 Darmoni, S. J., Leroy, J. P., Thirion, B., Douyère, M., Lacoste, B., Godard, C., Rigolle, I., Brisou, M., Videau, S., Goupy, E., Piot, J., Quéré, M., Ouazir, S. & Abdulrab, H. A search tool based on ‘encapsulated’ MeSH thesaurus to retrieve quality health resources on the Internet. *Medical Informatics and the Internet in Medicine* 2001, **26**, 165–78.
- 18 Soualmia, L. F., Barry, C. & Darmoni, S. J. Knowledge-based query expansion over a medical terminology oriented ontology. In: Dojat, M., Keravnou E. & Barahona, P. (eds). *Springer's Lecture Notes of Computer Science. Artificial Intelligence in Medicine. Proceedings of the 9th Conference on Artificial Intelligence in Medicine in Europe, AIME* 2003, 209–13.
- 19 Zweigenbaum, P., Baud, R., Burgun, A., Namer, F., Jarrousse, E., Grabar, N., Ruch, P., Le Duff, F., Thirion, B. & Darmoni, S. J. Towards a unified medical lexicon for French. In: Surjan, G., Engelbrecht, R. & McNair, P. (eds). *Proceedings of MIE 2003, Eighteenth International Congress of the European Federation for Medical Informatics*. Amsterdam: IOS Press Publisher, Stud Health Technol Inform, 2003, **95**, 415–20.
- 20 Horrocks, I., Patel-Schneider, P. F., van Harmelen, F. From SHIQ to OWL: the making of a web ontology language. *Journal of Web Semantics* 2003, **1**, 7–26.

Key Messages

Policy

- Enhancing the MeSH thesaurus is a necessity to index and retrieve health resources instead of scientific article.
- Quality-controlled health gateways are mandatory to apply quality services measures to support systematic resource discovery.

Practice

- To enhance the MeSH thesaurus, the CISMef team has introduced two new concepts: metaterms and resource types.
- Resource types are a generalization of Medline publication types.
- A metaterm is a medical specialty which has semantic links with MeSH keywords, qualifiers and resource types.