Affiliation of a resource type to a MeSH term in a quality-controlled health gateway

Stéfan J. Darmoni[a], Benoit Thirion[a], Filip Ionut-Florea[a], Alexandrina Rogazan[a], Catherine Letord[a], Gaétan Kerdelhué[a], Jean Nicolas Dacher[a]

[a] *CISMeF, Rouen.University Hospital & GCSIS, LITIS EA 4051, Institute of Biomedical Research, University of Rouen, France*

Abstract

Background*: CISMeF ([French] acronym for Catalog and Index of French Language Health Resources on the Internet) is a quality-controlled health gateway to catalog and index the most important and quality-controlled sources of institutional health information in French. CISMeF uses the MeSH thesaurus as a standard tool for organizing information. The CISMeF team enhanced the MeSH thesaurus with the introduction of two new concepts, respectively resource types and metaterms.* ***Objective****: The goal of this article is to describe and to evaluate another new enhancement of the CISMeF terminology: affiliation of a resource type to a MeSH term or to a MeSH (term/subheading) pair.* ***Methods****: CISMeF resource types (RT) are organized similarly to MeSH terms and subheadings (SH), in a hierarchical structure allowing the explode property and a maximum of five-level depth. In a sample of 15 requests with the most frequent diseases in the CISMEF catalogue, we compared the precision of the affiliation of resource types with the affiliation of the most frequently employed subheading in the CISMeF catalogue therapy.* ***Results****: The number of resources with at least one affiliation of a RT is 412 (2.8%) in the overall CISMeF catalogue. This figure is significantly lower than the number of resources with at least one affiliation of a SH (N= 8,110; 55.1%; p <0.0001; Mac Nemar's test). A significant difference was also present in the evaluated sample between the number of resources with at least one affiliation of a RT vs. the number of resources with at least one affiliation of a SH (39/2,019 (1.9%) vs. 1,001/2,019 (49.6%); p <0.0001; Mac Nemar's test). In our sample, a request with an affiliated RT is nine times more precise than the equivalent request with a floating RT.* ***Conclusion****: Affiliation of a RT to a MeSH (term/subheading) to create a triplet allows a better precision of the information retrieval in a health gateway.*

Keywords

Abstracting and Indexing; Cataloguing; Controlled Vocabulary; France; Information Storage and Retrieval; Internet; Medical Subject Headings; Publication types; Subheadings;.

Introduction

The Internet and in particular the Web has become an extensive health information repository. In this context, several quality-controlled health gateways have been developed [1]. Quality-controlled subject gateways were defined by Koch [2] as Internet services which 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.

Among several quality-controlled health gateways, CISMeF ([French] acronym for Catalog and Index of French Language Health Resources on the Internet) was designed to catalog and index 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 (N=14,714). CISMeF is manually indexed by a team of four indexers, which are medical librarians and systematically checked by the chief information scientist (the "super-indexer"). Its URLs are http://www.chu-rouen.fr/cismef or http://www.cismef.org.

CISMeF uses two standard tools for organizing information: the MeSH thesaurus [3] and several metadata element sets, in particular the Dublin Core metadata format (URL:http://www.dublincore.org) [4]. The use of specific metainformation is crucial in order to improve the recall and precision of internet searches [6]. As proposed by Hoelzer et coll. [6], CISMeF uses XML and RDF to meet these requirements.

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 (RT) and metaterms (MT), which have been previously described in [1]. The CISMeF terminology is shown in Figure 1. CISMeF resource types (RT) are an extension of the publication types of MEDLINE.

As defined by the Dublin Core Metadata Initiative (URL: http://www.dublincore.org/documents/dcmi-terms/) [4], a RT

is used to categorize the nature or genre of the content of the resource. MeSH (term/subheading) pairs describe the topic of the resource. RT is one of the fifteen Dublin Core repeatable and optional elements [4]. A metaterm is generally a medical specialty or a biological science, which has semantic links with one or more MeSH terms, subheadings and RTs. To construct a taxonomy of medicine, the publishing division of American Medical Association (AMA) took as its precedent the simplified access to MeSH via CISMeF metaterms [5].
However, the MeSH thesaurus was originally intended to index scientific articles for the MEDLINE database. In order to customize it to the broader field of health Internet resources, we have been developing several enhancements [1] to the MeSH since 2000, including the generalization of Major/Minor weights to RT and MT.
The goal of this article is to describe and to evaluate another new enhancement of the CISMeF terminology: affiliation of a RT to a MeSH term or to a MeSH (term/subheading) pair. The main idea is to obtain a new affiliation concept [MeSH (term/subheading)\CISMeF RT] constituting a triplet, in order to be more precise during the manual indexing process and therefore during the information retrieval process. Affiliation of a RT is similar conceptually to the affiliation of SH, taking into account the respective definitions of a RT and a SH (see above). Three PhD students of the CISMeF team were involved in this work, which relates to the optimization of information retrieval in the Doc'CISMeF search engine, as well as to the automatic indexing of textual resources in CISMeF, and to the automatic indexing of bimodal (text + image) resources in CISMeF

Materials and Methods

Background

Compared to the publication types of MEDLINE, the CISMeF RTs are more diverse, with specific RTs dedicated to electronic health resources, such as *association*, *patient information*, *community networks*, or *clinical guidelines*. For example, in the case of a clinical guideline about carbon monoxide intoxication, 'carbon monoxide poisoning' is the MeSH term and 'clinical guidelines' is the resource type. CISMeF RTs are organized similarly to MeSH terms and subheadings, in a hierarchical structure with subsumption relationships (allowing the explode property) and a maximum of five-level depth. The MEDLINE publication types were mainly a flat list till 2005 (see URL: http://www.nlm.nih.gov/mesh/pubtypes2005.html). Since 2006, MEDLINE publication types has also a hierarchical structure.
In preparation of the PhD thesis aiming at the automatic indexing of bimodal resources (text + image) in CISMeF, we have introduced new resources types, which are all medical image types. The medical image types list (n=112) are in majority MeSH terms from the *diagnostic imaging* term and its sub-tree composed of narrower terms. The overall number of resources types in December 2006 is 257. The controlled list of is available at the following URL: http://www.chu-rouen.fr/documed/typeeng.html. This RT list was manually built and maintained by the CISMeF team since 1997.
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 maximize the recall then the search answers (which means the Doc'CISMeF search ORes the answers for the MeSH term and the answers for the RT) when the request 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).
The medical image types list has been reviewed and updated manually by one medical imaging expert (JND). Medical image types are used in the CISMeF database to index a health resource containing images. For example, a teaching resource about choledocholithiasis, which includes ultrasonography images, will be indexed with the RTs *ultrasonography* and *teaching material*.
If this teaching resource contains a paragraph describing the ultrasonography of the choledocholithiasis, the librarian will use the MeSH (term/subheading) pair (*choledocholithiasis*/*ultrasonography*). If necessary, two others subheadings can be used: '*radiography*' and '*radionuclide imaging*'.
In the case of other medical image types, MeSH terms will be used. For example, if the resource about choledocholithiasis includes text about magnetic resonance imaging, it will be indexed with the two MeSH terms '*choledocholithiasis*' and '*magnetic resonance imaging*'. If there is an image of magnetic resonance imaging, it will be indexed with the RT '*magnetic resonance imaging*'.

Affiliation

Until the introduction of images types, we did not think to use the existing RTs for affiliation to a specific aspect of MeSH term or (term/subheading) pairs. In April 2004, the creation of medical image types led us to propose a refinement of the CISMeF terminology and thus a refinement of manual indexing procedures: for a number of specific RTs, mainly images types but not exclusively (e.g. *multiple choice quiz* also applies here), we proposed to affiliate a RT to a MeSH term or to a MeSH (term/subheading) pair. Thus we can obtain a [MeSH (term/subheading)\CISMeF RT] triplet, where the backslash character '\' represents the RT affiliation (the slash character / represents in MEDLINE the affiliation of subheading to a term). This approach can be viewed as an extension of the affiliation of a subheading to a MeSH term.
Therefore, a teaching resource about choledocholithiasis, which includes ultrasonography images, should be indexed with the (term\RT) pair (choledocholithiasis\ultrasonography). If the teaching resource specifies for example that ultrasonography images are valuable for the diagnosis of choledocholithiasis, the [MeSH (term/subheading)\CISMeF RT] triplet [(choledocholithiasis/diagnosis)\ultrasonography] should be used. Another triplet can be viewed in the following description of a resource indexed in the CISMeF catalogue; the triplet "***atrial fibrillation/diagnosis\electrocardiography***" indicates that this resource contains an EKG image to contribute to the diagnosis of atrial fibrillation.
Atrial fibrillation (The) –
French pre-residency program examination : question 236 –

Pr Vanzetto G (Grenoble University), M. Defaye P. (Grenoble University)
[publisher : Joseph Fourier University, Medical School ; definition, etiology, physiopathology, anatomical sequels, diagnosis, evolution and prognosis, treatment ; printable version, references, pre-necessary, exercises ; language : French ; format : xml ; access : free ; date : 2005 ; visited on November 2006]. Grenoble-Fr
*keywords : cardiology / education ; *atrial fibrillation ;* ***atrial fibrillation / diagnosis \ electrocardiography*** *; atrial fibrillation / etiology ; atrial fibrillation / physiopathology ; atrial fibrillation / therapy ; prognosis ; signs and symptoms*
*type : *educational courses ; electrocardiography ; *multiple choice quizz*
Since April 2004, all the CISMeF resources indexed with an image RT (N=1288 out of 14,714; 8.8%) have been reviewed by the five CISMeF medical librarians to check the need of any affiliation of a RT.

Evaluation

To evaluate the precision of the affiliation of RTs, we have selected the 15 diseases from the C and F MeSH trees which are the most frequently used for the indexing in the CISMeF catalogue (see Table 1). For the RT, we have chosen the top level hierarchy "image" because it is the most used RT for affiliation. For each MeSH term, two requests were performed: one request with floating RT (MeSH term AND image) and one request with affiliated RT (MeSH term\image).
We compared the precision of the affiliation of RTs with the affiliation of subheadings in the same catalogue. We have used the most frequently employed subheading in the CISMeF catalogue: therapy (N=4,267, 29.0% of the catalogue).

Table 1 - Evaluation of request using affiliated vs. floated resources type & subheading

MeSH terms (Number of resources included in CISMeF)	Floated resource type image[1]	Affiliated resource type image[2]	Floated subheading therapy[3]	Affiliated subheading therapy[4]
Pain (287)	63	10	192	150
AIDS (270)	23	3	154	101
Diabetes mellitus (207)	38	3	126	79
Cross infection (202)	27	1	172	160
Breast neoplasms (161)	37	7	115	105
Hypertension (133)	34	2	96	63
Alcoholism (110)	7	1	66	40
Tuberculosis (106)	18	5	76	61
Hepatitis C (104)	7	0	70	50
Hepatitis B (81)	4	0	68	54
Obesity (80)	21	1	60	34
Neoplasm metastasis (76)	35	1	70	25
Lung neoplasms (70)	15	4	38	25
Prostate neoplasms (69)	13	1	50	41
Blindness (63)	2	0	15	10
Total (2,019)	344 (17.%; 44/2,019)	39 (11.3%; 39/344)	1,368 (67.8%; 1,368/2.019)	1,001 (73.2%1.001 /1,368)

[1] Request: MeSH term AND image
[2] Request: MeSH term\image
[3] Request: MeSH term AND therapy
[4] Request: MeSH term/therapy

Results

The number of resources with at least one affiliation of a RT is 412 (2.8%) in the overall CISMeF catalogue. This figure is significantly lower than the number of resources with at least one affiliation of a SH (N= 8,110; 55.1%; p <0.0001; Mac Nemar's test).
A significant difference was also present in the evaluated sample (see Table 1) between the number of resources with at least one affiliation of a RT vs. the number of resources with at least one affiliation of a SH (39/2,019 (1.9%) vs. 1,001/2,019 (49.6%); p <0.0001; Mac Nemar's test).
The number of resources with at least one affiliation of a RT (and of a SH) in the last 500 resources included in CISMeF is higher when compared to the overall catalogue. (173/500; 34.6% for RT vs. 415/500; 83.0% for SH). Nonetheless, in the sample of the last 500 resources included in CISMeF, there is still a significant difference between the number of resources with at least one affiliation of a RT vs. the number of resources with at least one affiliation of a SH (p <0.0001; Mac Nemar's test).
Furthermore, the average use of RTs per resource is also significantly lower than the average use of SHs in the overall CISMeF catalogue (1.95 vs. 4.47, p <0.0001; paired Student's T test). This result is correlated by the significant difference between the number of resources with the floating RT image (Request: MeSH term AND image) vs. the number of resources with the floating subheading therapy (Request: MeSH term AND therapy) in the evaluated sample: 344 (17.0%;

344/2,019) vs. 1,368 (67.8%; 1,368/2,019) (p <0.0001; Mac Nemar's test) (see Table 2).

In the evaluated sample, to measure the precision of the affiliation of RTs, we compared the number of resources affiliated with at least one RT of the RT tree "image" vs. the number of resources with a floating RT image (Request: MeSH term AND RT image). The ratio is 39/344 (11.3%). Therefore, in our sample, a request with an affiliated RT is nine times more precise than the equivalent request with a floating RT.

Then, to measure the precision of the affiliation of SHs, we compared the number of resources affiliated with the SH "therapy" vs. the number of resources with the floating SH "therapy" (Request: MeSH term AND SH therapy). The ratio was 1,001/1,368 (73.2%). In our sample this means that a request with an affiliated SH is only 1.3 times more precise than the equivalent request with the floating SH.

The possibility to affiliate RTs has been implemented in the Doc'CISMeF search engine in every search mode (Simple, Advanced, Boolean, and Step by Step). Nonetheless, by default, this improvement will not modify the information retrieval process in the Doc'CISMeF search engine (URL: http://doccismef.chu-rouen.fr), in the Simple Search mode. As described in , this information retrieval process is based by default on implicit query processing and the algorithm is based on maximizing the recall. Implicit query processing means that the end-user does not interact with the system to improve the quality of the information retrieval. In that case, the search engine tries to maximize the mapping of the end-user's request to the CISMeF terminology. This CISMeF mapping algorithm [1] has similarities with PubMed's Automatic Term Mapping [7].

For example, an ambiguous query such as "Diagnosis of choledocholithiasis with ultrasonography", will be transformed into the following Boolean query in order to provide the user with the wider range of documents likely to meet their needs, using MeSH terms, floating SHs and floating RTs: "diagnosis AND choledocholithiasis AND ultrasonography". Nonetheless, a trained user may use the affiliation of RTs (combined or not with the affiliation of SH) to obtain a more precise retrieval process. In the last example, the user may enter the triplet "(choledocholithiasis/diagnosis)\ultrasonography" to be as precise as possible if s/he is searching for an image of ultrasonography to diagnose choledocholithiasis.

To lead CISMeF users to use this refinement among others enhancements of the CISMeF terminology [1], two training sessions are organized on a monthly basis by the CISMeF team, targeting mainly medical students, as well as librarians and health professionals. How to use the affiliation of RT will part of the training session as it is already the case for affiliation of a subheading.

Discussion

The goal of this article was designed to evaluate in terms of precision the affiliation of a RT to a MeSH term or to a MeSH (term/subheading) pair. This study is also a partial answer to a comparative study performed by Abad Garcia et coll. [8] among six European health gateways: Organizing Medical Networked Information (OMNI, now Intute Life Sciences), (URL: http://www.intute.ac.uk/healthandlifesciences/), Diseases, Disorders and Related Topics (DDRT) (URL: http://www.mic.ki.se/Diseases/), Health on the Net Foundation (http://www.hon.ch) [9], with two tools: Medical Document Hunter (MedHunt) and the Multilingual and Intelligent Search Tool Integrating Heterogeneous Web Resources (HONSelect) search engine, Computer Aid Learning (CAL) Reviews and CISMeF. Although CISMeF was rated second after OMNI, CISMeF has been criticized because "failure on precision may be due to exhaustive indexing" [8]. The implementation in the CISMeF health gateway of the affiliation of RT is now optimizing the precision.

In the evaluated sample, a request with an affiliated RT is nine times more precise than the equivalent request with a floating RT, while a request with an affiliated SH is only 1.3 times more precise than the equivalent request with the floating SH. This significant difference is partially correlated with the average use of RTs per resource, which is significantly lower than the average use of SHs in the overall CISMeF catalogue (1.95 vs. 4.47). The respective precision of the RT affiliation vs. SH affiliation may decrease with time as the figures are significantly different for the last 500 resources included in the CISMeF catalogue. That is why this evaluation should be performed again in one year's time.

To affiliate RT to a MeSH term (or to a MeSH term/subheading pair) has one drawback: the retrieval of the Doc'CISMeF search engine leads to the whole document and not the retrieval of the specific image(s) that lead to the indexing of the affiliated RT. Nonetheless, this drawback is similar with the MEDLINE/PubMed Website (URL: http://www.pubmed.gov) and the affiliation of a subheading to a MeSH term, which also lead to the retrieval of an article and not to the retrieval of the specific paragraph(s) that lead to the indexing of the affiliated SH.

The triplet MeSH term/SH\TR may also have (also) a teaching value for medical students and health professionals; e.g. breast neoplasms/prevention & control\mammography should be interpreted as follows: mammography has an important role in the prevention of breast neoplasms.

As metaterms for AMA [5], the comprehensive list of RT could be used by several health institutions, specially those which are Dublin Core compliant (see URL: http://www.chu-rouen.fr/documed/dc.html). Affiliation of RT and the various enhancements of the MeSH thesaurus described in [1] could be applied to any Web site that uses the MeSH thesaurus, firstly the MEDLINE/Pubmed bibliographic database, with the restriction of using Publication Types instead of RT, and also health gateways such as Intute Life Sciences, DDRT, Health on the Net, and the US National Guideline Clearinghouses (URL: http://www.guideline.gov) [10].

We are currently preparing the Cogni-CISMeF project, which aims at initiating a human computer interaction in order to disambiguate subject queries [7], such as the one seen in the previous example "Diagnosis of choledocholithiasis with ultrasonography". Cogni-CISMeF would prompt the user for context information: Do you require a paragraph or an image? If the end-user chooses the image, the query will be restricted to [(choledocholithiasis/diagnosis)\ultrasonography].

We are also working on adapting the algorithm developed by the CISMeF team to automatically index MeSH (keyword/qualifier) pair [11]. The idea is to extend this algorithm to the automatic indexing of [MeSH (keyword/qualifier)\ CISMeF RT] triplet. The first aim of the bimodal indexing tool is to automatically recognize six RTs, which are the main six modalities used in medical imaging: standard radiography, arteriography, ultrasonography, radionuclide imaging, CT scanner and magnetic resonance imaging.
Finally, we are also collaborating with a PhD student of the Rennes Laboratory of Medical Informatics who is also working on the automatic indexing of bimodal resources (text + images) but in a different context: electronic patient record instead of digital library (health resources on the Internet). Therefore, the list of resources types will not be MeSH-oriented but CCAM-oriented [12], a coding system for procedures in France.

Conclusion

Affiliation of a RT to a MeSH (term/subheading) to create a triplet allows a better precision of the information retrieval in a quality controlled health gateway

References

[1]. Douyère M, Soualmia LF, Névéol A, Rogozan A, Dahamna B, Leroy JP, Thirion B, Darmoni SJ: Enhancing the MeSH thesaurus to retrieve French online health resources in a quality-controlled gateway. Health Info Libr J 2004 Dec: 21(4):253-61.

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

[3]. Nelson SJ, Johnson WD, Humphreys BL: Relationships in Medical Subject Headings in Relationships in the organization of knowledge. In Bean and Green, eds. Kluwer Academic Publishers, 2001;pp. 171-84.

[4]. Dekkers M, Weibel S: State of the Dublin Core Metadata Initiative. D-Lib Magazine April 2003: vol 9. Number 40.

[5]. Mc Gregor B. Constructing a concise medical taxonomy. J Med Libr Assoc. 2005 January; 93(1): 121–123.

[6]. Hoelzer S, Schweiger RK, Boettcher H, Rieger J, Dudeck J. Indexing of Internet resources in order to improve the provision of problem-relevant medical information. Stud Health Technol Inform. 2002;90:174-7.

[7]. Smith AM: An examination of PubMed's ability to disambiguate subject queries and journal title queries. J Med Libr Assoc 2004 Jan: 92 (1): 97-100.

[8]. Abad Garcia F, Gonzalez Teruel A, Bayo Calduch P, de Ramon Frias R, Castillo Blasco L. A comparative study of six European databases of medically oriented Web resources. J Med Libr Assoc. 2005 Oct;93(4):467-79.

[9]. Boyer C, Geissbuhler A. A decade devoted to improving online health information quality. Stud Health Technol Inform. 2005;116:891-6.

[10]. Web site offers database of national guidelines. Healthc Benchmarks. 1999 Mar;6(3):30-1.

[11]. Névéol A, Rogozan A, Darmoni SJ: Automatic indexing of health resources in French with a controlled vocabulary for the CISMeF catalogue: a preliminary study. In Eleventh World Congress on Health and Medical Informatics, MedInfo 2004:(CD)1772.

[12]. Trombert-Paviot B, Rodrigues JM, Rogers JE, Baud R, Van Der Haring E, Rassinoux AM, Abrial V, Clavel L, Idir H.:GALEN: a third generation terminology tool to support a multipurpose national coding system for surgical procedures. Int J Med Inf 2000 Sep: 58-9:71-85

Address for correspondence

SJ. Darmoni, CISMeF, Rouen.University Hospital, 1 rue de Germont, 76031 Rouen Cedex, France & GCSIS, LITIS EA 4051, Institute of Biomedical Research, University of Rouen, France