MEDINFO 2007
HISA Ltd Australia
© 2007 The authors. All rights reserved.

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*

Introduction

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 and several metadata element sets, in particular the Dublin Core metadata format (URL:http://www.dublincore.org). The use of specific metainformation is crucial in order to improve the recall and precision of internet searches As proposed by Hoelzer et coll., 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). 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/), 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.

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 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).

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. The overall number of resources types in December 2006 is 257.

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

Affiliation

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. 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. 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).

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). 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 has similarities with PubMed's Automatic Term Mapping.

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, 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.

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

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