

Translating MeSH Concepts.

Lina Soualmia, Catherine Letord, Tayeb Merabti, Nicolas Griffon, Jacques Manel, Stéfan Darmoni

► **To cite this version:**

Lina Soualmia, Catherine Letord, Tayeb Merabti, Nicolas Griffon, Jacques Manel, et al.. Translating MeSH Concepts.. Studies in Health Technology and Informatics, IOS Press, 2013, 192, pp.1007. <inserm-00854292>

HAL Id: inserm-00854292

<http://www.hal.inserm.fr/inserm-00854292>

Submitted on 27 Aug 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Translating MeSH Concepts

Lina F. Soualmia^{a,b}, Catherine Letord^a, Tayeb Merabti^a, Nicolas Griffon^a,
Jacques Manel^c, Stéfan J. Darmoni^{a,b}

^a LITIS-TIBS EA 4108 & CISMef, Rouen University Hospital, Rouen, France.

^b INSERM, UMR_S 872, ICS, équipe 20, Paris, France.

^c Poisons and Toxicovigilance Centre, Nancy University Hospital, Nancy, France.

Abstract and Objective

The concept-oriented structure of the MeSH® thesaurus is not yet in common use. Nevertheless, it has been shown that a concept-based querying of PubMed may be of interest. To take full advantage of the concept-oriented structure of MeSH in the information retrieval tool associated with the CISMef catalogue, it was necessary to translate such concepts into French.

Keywords:

MeSH Headings; Translations; Controlled Vocabularies;

Introduction

A recent study [1] has demonstrated the interest of using MeSH Concepts for improving information retrieval. To index documents with MeSH Concepts instead of MeSH Descriptors or MeSH Supplementary Concepts, MeSH Concepts need to be translated into French to be used in the CISMef catalogue.

Material and Methods

In addition to "broader", "narrower" and "related" relations, certain MeSH Concepts may have other attributes: a Registry Number (RN) or a Related Registry Number (RRN) which may correspond to the Chemical Abstracts Service (CAS) Registry Numbers. Each MeSH Concept may have a RN. In this case, the RN is the CAS Number of the Concept itself. Only MeSH Concepts that are the preferred concept for a MeSH Descriptor or a MeSH Supplementary Concept may have a RRN. In this case, the RRN is the CAS Number of a subordinate Concept. Many MeSH Concepts (n=81,255; 25.1%) have a CAS Number. The CISMef team has also integrated into the HeTOP [2] a French terminology on main toxic substances for humans (BNPC French acronym). It contains 91,750 terms in French.

Mapping MeSH Concepts and BNPC terms

The most straightforward method consisted in a simple join between MeSH Concept and BNPC term using CAS Number. It was then possible to automatically translate this MeSH Concept into French and the corresponding BNPC term into English. When RRN is involved, a "see also" relation between two terms from two different terminologies: the MeSH Descriptor or the MeSH Supplementary Concept, which has a RRN (CAS Number) and the BNPC term with the same CAS Number.

Simple Natural Language Processing Method

The normalization program ("Norm") which is provided with the Lexical Tools of the UMLS was used in this second method. The MeSH Concept Labels in English were normalized and compared to all the other normalized terms in the HeTOP portal. When an exact match was found, the translation of the French target term was proposed as one possible translation of the MeSH Concept Label.

Results

Thanks to the mappings based on the CAS Number, 13,440 MeSH Concepts were translated into French. These MeSH Concepts might be:

- preferred concept for MeSH Descriptors for 2,046 of them (7.7%),
- preferred concept for MeSH Supplementary Concept for 9,174 of them (4.5%),
- non-preferred or subordinate concept for MeSH Supplementary Concept or MeSH Descriptor for 2,220 of them (2.4%) and the creation of 2,220 "see also" relationships between a BNPC term and a MeSH Descriptor or MeSH Supplementary Concept.

Overall, 13,440 BNPC terms were also translated into English. The NLP method allowed the translation of 8,701 MeSH Concepts: 3,460 preferred concepts for MeSH Descriptors, 3,455 preferred concepts for MeSH Supplementary Concept and 1,786 non-preferred concepts.

As 52,996 MeSH Concepts had already been manually translated before this study there are now 65,621 MeSH Concepts out of 324,191 (20.2%), which have been translated into French. Moreover, 9,516 new synonyms have been added.

References

- [1] Darmoni SJ, Soualmia LF, Letord C, et al. Improving information retrieval using MeSH Concepts: a test case on rare and chronic diseases. *Journal of Medical Library Association* 2012; 100(3): 176–183.
- [2] Merabti T, Soualmia LF, Grosjean J, Joubert M, Darmoni SJ : Aligning biomedical terminologies in French: towards semantic interoperability in medical applications. Chapter in *Book Medical Informatics, InTech*, 2012, 41–68.