A combined onomasiological and semasiological approach to the development of a comprehensive interface terminology for information storage and retrieval in primary health care

Robert Vander Stichele, Marc Jamoulle, Elena Cardillo, Joseph Roumier, Laurent Romary, Erik Van Nimwegen, Maxime Warnier

A combined onomasiological and semasiological approach to the development of a comprehensive interface terminology for information storage and retrieval in primary health care.

Robert Vander Stichele, Marc Jamoulle, Elena Cardillo, Joseph Roumier, Laurent Romary, Dirk Van Nimwegen, Maxime Warnier, CISP-CLUB, Bruxelles, October 15, 2012

Since 2011, the Meriterm (Medical End-user Reference Interface Terminology) Group has started the development of a reference terminology for Primary Health Care professionals and patients.
It is a multidisciplinary group with the participation of information scientists, medical doctors, linguists, knowledge engineers, IT-specialists and sociologists.
The Group is a consortium between the University of Ghent, including the Heymans Institute of Pharmacology and the Language and Technology Department (LT3), Belgium (Flanders); the Centre of Excellence For Technology in Information and Communication (CETIC, Belgium, (Wallonia); and the Fondazione Bruno Kessler (FBK), Italy (Trento), with the participation of experts from IRSS, University of Louvain, Belgium and from INRIA, France"

An elaborated approach to creating a reference terminology has been developed.
First, terms are selected from everyday interactions between doctors (occurrences in medical records, in guidelines), based on frequency count and relevance.
Then an onomasological approach is applied to make a choice on basic concepts most intricately related to the most common sense of the selected terms. If possible a perfect concordance for that concept is looked for in the SNOMED nomenclature, and if not possible, in UMLS nomenclature, or genuinely defined. The nature of the mapping from the concept in the reference terminology to external international classification is qualitatively rated (exact, nearly exact, match to higher or lower level of granularity). In addition, each concept in the Reference Terminology (RT) is also mapped to classifications such as ICPC, ICD. Mapping to thesauri (MeSH) and other nomenclatures (LOINC, NANDA) is foreseen.

The management of the collection of concepts in the RT and the mapings to nomenclatures, thesauri and classifications is managed in a multilingual TMF (Terminological Markup Format, ISO 16642) application. For each participating language, a literal translation of the selected (and considered preferred) lexical representation in English (as a meta-language) is given, as well as possibly a more suitable clinical term, and a lay language term. In addition, links to similar terms or synonyms for each concept will be provided in an unilingual extention of the TMF application for each language.
In a combined semasiological approach, the concepts of this RT will also be linked to a number of unilingual LMF (Lexical Markup Format, ISO 24613) applications, one for each language, to manage the potential other senses of their lexical representations. At this level, also subtle differences between related languages (e.g. Portuguese in Portugal and in Brazil, English in the UK and in the US, Dutch in The Netherlands and in Belgium). This will enable also linking to Natural Language Processing resources, such as WordNet.
A deliberate choice was made not to introduce a new ontology into the collection of basic applications, but to use the ontologies behind the NLP systems and the international nomenclatures and classifications to which links and mappings were made.

All TMF and LMF applications in this project will be structured in close compliance to the corresponding ISO standards and relate to the ISOCat directory for data categories and conceptual domains (www.isocat.org). Through an export of the data in OWL/RDF, the data will be published as Linked Open Data on the semantic web and accessible to SPARQL queries. The TMF and LMF applications will be available for on-line maintenance with an interactive Semantic MediaWiki application. Under a Creative Commons License, the data will be available for computer applications for terminology, sustaining medical registration,

information retrieval, and scientific analysis of medical data. In addition, the data may support communication between Electronic Health Records, kept by physicians and Patient Health Records, kept by patients.

The Meriterm Group has recently been awarded with a scientific prize by the Monnet Challenge organized by the European Union project Monnet dedicated to linguistic resources.