Cross-lingual access to biomedical terminologies and ontologies

Julien Grosjean[1], Lina F. Soualmia[1], Tayeb Merabti[1], Nicolas Griffon[1], Badisse Dahamna[1], Stéfan J. Darmoni[1]

CISMeF, Rouen University Hospital & TIBS, LITIS EA 4108, Institute of Biomedical Research, University of Rouen, France

{julien.grosjean, lina.soualmia, tayeb.merabti, nicolas.griffon, badisse.dahamna, Stefan.darmoni}@chu-rouen.fr

Abstract. The Health Terminology/Ontology Portal (HeTOP) is a repository dedicated to health professionals and students. It provides access to 45 health knowledge bases (including terminologies and ontologies) available in 23 different languages. Several methods and technologies have been developed to create this portal, dedicated to both human and computers. HeTOP is a valuable tool for a wide range of applications and users, especially in education and resource indexing but also in information retrieval or performing audits in terminology management. A total of 5,355,000 terms and 4,890,000 relations are included in HeTOP. To our knowledge, this kind of multi-terminology and cross-lingual portal is the first of its kind. Non-European languages have been integrated recently. The conceptual approach used in the model allows integration of any language while maintaining valid relations between concepts.

Keywords. Controlled Vocabulary; Internet; Information Storage and Retrieval; Terminology as subject; Health terminologies and ontologies; Cross-lingual access to health information; Portal.

1 Introduction

Terminologies and ontologies (T/O) are not only increasingly more complex with rich semantic relations, but also more diversified and exploited. Rich semantic relations are defined as relations that provide the end-user with an added-value besides the classic broader-narrower (BT-NT). These knowledge resources are mainly used to index (or annotate) or perform complex tasks such as ontology reasoning.

A terminology server has been defined as a tool to manage and to give access to various terminologies [1]. Several terminology servers have already been developed, mostly in English, in particular BioPortal [2], developed by the US National Center for Biomedical Ontologies (NCBO). Since 2006, the CISMeF team has been developing a terminology portal which originally focused on French T/O (URL: http://pts.chu-rouen.fr) [3].

The objective of this paper is to demonstrate the various interests of a cross-lingual terminology portal, including (a) to index any document in a multi-terminology cross-lingual mode, (b) to teach rare diseases or anatomy, (c) to develop multi-terminology automatic indexing and information retrieval tools and (d) to perform audits in terminology management.

2 Methods

Dealing with these kinds of T/O is not an easy task due to structure, size, nature and specificity. First, a meta-model was created in order to integrate all T/O and one global generic system. This meta-model has been validated as we managed to integrate not only any terminology into it, but also ontologies [4]. The meta-model is compliant with the ISO 25964-1 data model and the ISO 25964-2 norm for interoperability.

HeTOP terminologies are implemented as light Ontology Web Language (OWL) ontologies. On the other hand, moving from an ontological to a terminological representation is based on a reification process. In this way, formal ontologies are "degraded" to fit this multi-terminology model. This meta-model is basically cross-lingual because preferred terms, synonyms or other textual attributes can be defined by a language code (en for English, fr for French, etc.). Each terminology T of HeTOP is built as an enrichment of this model. We have combined different data sources for each available terminology language (UMLS, official national sources of ICD-10, etc.). The resulting physical model is a powerful support to perform huge data handling (integration and information retrieval). To ensure interoperability between T/O, Natural Language Processing tools have been developed and validated [5]. Finally, a web application was developed, dedicated to both health professionals and students, with an adapted graphic interface and an efficient search engine. A French InfoButton providing access to around 50 Web knowledge bases such as PubMed [6] has been integrated into the HeTOP [7].

This tool was evaluated by two consecutive groups of second year medical students in September 2010 and September 2011 respectively.

3 Results

A total of 45 terminologies are included in HeTOP, with 1,570,000 concepts, 3,680,000 synonyms, 192,000 definitions and 4,890,000 relations. A total of 23 different languages are available. Some of these European languages do not use the Latin alphabet e.g. Greek or Russian and some non-European languages were also introduced into HeTOP such as Arabic or Japanese).

Thirty two of these terminologies are not yet included in the UMLS; among them, some are developed by the World Health Organization (e.g. ATC for drugs).

In the current cross-lingual version, it is possible to navigate both as a matrix between the 45 T/O and in the mean time between 23 languages.

To consult the crosslingual HeTOP (http://www.hetop.eu/, click on “Se connecter”; login=fmauser, password=fmapass). HeTOP freely provides access to many terminologies such as ICD10 and FMA in several languages.

Enrichment of T/O. The CISMeF has manually translated several T/O.~~;~~ For the MeSH thesaurus, the CISMeF team has added: 22,039 synonyms to the MeSH Descriptors, 163 synonyms to the MeSH Qualifiers. It has also manually translated 20,909 MeSH Supplementary Concepts (10.17%) and has added to them a set of 6,974 synonyms in French.

Ontology auditing. For each Orphanet disease which has a semantic exact match with OMIM, it is now possible for the ontologist to confront the Orphanet phenotypes to the HPO phenotype. For example, for the Marfan disease, Orphanet provides a semantic link to 65 signs, when HPO provides 51 signs. It is then easy for ontologists from HPO and Orphanet to review the discrepancies between the two ontologies.

Teaching. HeTOP was used as a tool for teaching rare diseases and anatomy to Rouen medical students since 2010. These two fields of medicine were chosen because one ontology already exists in anatomy (FMA) and two for rare diseases (Orphanet and HPO).

Evaluation. The results of the two qualitative evaluation surveys performed over the previous two years on two successive cohorts of Rouen Medical School students (second year) are as follows. The results of the questionnaire are displayed in Table 1.

Table 1. Results of the 2010 and 2011 evaluations of HeTOP.

	Mean (%) ± Std deviation 2010	Mean (%) ± Std deviation 2011	Mean (%) Overall
Interest in teaching	79.9 ± 12.9	87.5 ± 10.1	83.7
Design	57.2 ± 16.5	55.5 ± 19.2	56.35

4 Discussion

The HeTOP terminology portal presented here has the main functionalities of any terminology server. On one hand, HeTOP has several qualities:

The main added value of HeTOP when compared to NCBO Bioportal [2] or the EBI Ontology Lookup Service [8] is the possibility to access biomedical T/O using cross-lingual functionalities, allowing navigation among T/O and in the same time among languages. Moreover, the aim of HeTOP is not to be a simple repository of versioned T/O such as BioPortal: HeTOP is dedicated to end-users and offers resources of quality and some extra tools to help them to understand the T/O in order to use them in a proper way.

Another added value of HeTOP when compared to any UMLS browser [9] is the possibility it offers to access the main health terminologies in French or multi-lingual terminologies and the World Health Organization (WHO). To the best of our knowledge, the HeTOP is the first terminology portal with such specific emphasis on French T/O (more than 500,000 terms in French included in HeTOP).

Whereas assessment of HeTOP has demonstrated that its content was most appreciated by students, these studies show a need for improvement in its design. Indeed, this kind of portal is complex and necessitates further research on new user access. In addition, a wider study on the portal quality and its use would be of value.

HeTOP is mainly dedicated to medical librarians to index resources in a multi-terminology mode. HeTOP is also very useful not only for translators, terminologists, but also healthcare professionals, in particular physicians.

Acknowledgments.

This study was partially funded by several regional (Normandy), national (France) and European projects, respectively PlaIR (URL: www.plair.org), InterSTIS (URL: www.interstis.org) & PSIP (URL: www.psip-project.eu/).

References

1. Burgun A, Denier P, Bodenreider O, Botti G, Delamarre D, Pouliquen B, Oberlin P, Lévéque JM, Lukacs B, Kohler F, Fieschi M, Le Beux P. A Web terminology server using UMLS for the description of medical procedures. J Am Med Inform Assoc 1997 Sep-Oct: 4(5):356-363.
2. Noy N.F., Shah N.H., Whetzel P.L., Dai B., Dorf M., Griffith N., Jonquet C., Rubin D.L., Storey M.-A., Chute C.G., and Musen M.A. BioPortal: ontology and integrated data resources at the click of a mouse. Nucleic Acids Research, 2009, Page 1-4; Web Server Issue 10 http://nar.oxfordjournals.org/content/early/2009/05/29/nar.gkp440.full
3. Grosjean J, Merabti T, Dahamna B, Kergourlay I, Thirion B, Soualmia LF, Darmoni SJ. Health Multi-Terminology Portal: a semantics added-value for patient safety. Stud Health Technol Inform. 20111;66:129-38.
4. Golbreich C., Grosjean J. & Darmoni S.J. The FMA in OWL 2. AIME 2011, 204-214, Springer-Verlag.
5. Merabti, T; Soualmia, LF; Grosjean, J; Joubert, M & Darmoni, SJ. Aligning Biomedical Terminologies in French: Towards Semantic Interoperability in Medical Applications. Medical Informatics, Pages 41-68, InTech, 2012.
6. Griffon N, Chebil W, Rollin L, Kerdelhue G, Thirion B, Gehanno JF, Darmoni SJ. Performance evaluation of Unified Medical Language System®'s synonyms expansion to query PubMed. BMC Med Inform Decis Mak. 2012 Feb 29;12:12.
7. Darmoni SJ, Pereira S, Névéol A, Massari P, Dahamna B, Letord C, Kerdelhué G, Piot J, Derville A, Thirion B. French Infobutton: an academic and business perspective. AMIA Annu Symp Proc. 2008 Nov 6:920.
8. Cote R.G., Jones P., Apweiler R., Hermjakob H. The ontology lookup service, a lightweight cross-platform tool for controlled vocabulary queries. BMC Bioinformatics. 2006 Feb 28;7(1):97.
9. Bodenreider O. The Unified Medical Language System (UMLS): Integrating biomedical terminology. 2004, 32:267-270.