Health Multi-Terminology Portal

Julien Grosjean[1], Tayeb Merabti[1], Stefan J Darmoni[1]

1. CISMeF, Rouen University Hospital, Normandy & TIBS, LITIS EA 4108, Rouen University

Abstract— **Since the mid-90s, several quality-controlled health gateways were developed. In France, CISMeF is the leading health gateway to index Internet resources from the main institutions, using the MeSH thesaurus and the Dublin Core metadata element set. Since 2005, the CISMeF team has switched from a mono-terminological world (MeSH) to a multi-terminology universe, with the integration of 27 terminologies in its information system. The principal objective of this work was to create a Health Multi-Terminology Portal (HMTP) connected to the CISMeF Terminology Database to search concepts among all the health terminologies available in French (or in English and translated in French) included in this portal and to browse it dynamically.**

Keywords
Abstracting and indexing ; Cataloging ; Controlled Vocabulary ; Internet ; Database ; France ; Information Storage and Retrieval ; Internet ; Subject Headings ; Terminology as subject

I. INTRODUCTION

From 1995 to 2005, CISMeF [1] used two standard tools to describe and index the most important and quality-controlled sources of institutional health information in French : the MeSH thesaurus and its French translation by the French Medlars Center (French National Institute of Health), as well as several metadata element sets. Various enhancements of the MeSH thesaurus that the CISMeF team has been developing in order to adapt this terminology to the broader field of health Internet resources vs. scientific articles for the Medline bibliographic database where the MeSH thesaurus was originally built for, have previously been described [2]. Since 2005, the CISMeF team has undergone a major strategic shift : switching from a mono-terminological world to a multi-terminology universe for the overall CISMeF information system, which includes multi-terminology automatic indexing [3], multi-terminology information retrieval and integration of several terminologies (n=27) in the CISMeF terminology database. There is an increasing amount of interest today not only in developing and maintaining healthcare terminologies but also in making them interoperable within information technology (IT) systems delivering services to applications. Terminology server has been defined as a tool to manage and to give access to various terminologies [4].The principal aim of this work was to create an Health Multi-Terminology Portal (HMTP) largely inspired by the most recent advances [5] and connected to the CISMeF terminology database to search concepts among all the health terminologies available in French (or in English and translated in French) included in this portal and to browse it dynamically. The ultimate goal was to use this search : (a) to index resources manually or automatically in the CISMeF quality-controlled health gateway ; (b) to permit multi-terminology information retrieval ; (c) to evaluate the integrity of terminological data (audit) ; (d) to provide a new source of education for students. This project is a completely innovative work human-oriented to deal with terminologies ; this is the main difference between ontologies and terminologies and as data is more important for user (structure for computer), we decided to oversimplify ontologies and terminologies (if necessary) integrated in the HMTP but we are caring more to data and its representation for human beings.

II. MATERIAL AND METHODS

II.1. Integration of the terminologies

To integrate the terminologies in the CISMeF database (Oracle 11.1g database), three steps are necessary : (1) to design a terminology generic model into which each terminology model can be integrated ; (2) to design a process capable of integrating terminologies into the HMTP ; (3) to build and integrate inter-terminology mappings into the HMTP. Two inter-terminology mappings were performed : one based on UMLS concepts and one based on NLP tools developed by the CISMeF team [6]. A generic model was designed for the database in order to fit all the terminologies into one global structure : this database is the CISMeF terminology database. Then, a model of each terminology was designed as a specialization of the meta-model. The purpose of this generic model is to factorize the artefacts (classes, relationships, attributes) common to all the terminologies, thus facilitating integration of multiple terminologies within a single platform. Some artefacts, although specific to certain terminologies, must nevertheless be represented in order to avoid losing information outside the generic model. Consequently, a trade-off has to be selected in order to faithfully represent a terminology with no loss of information while removing artefacts shared by terminologies in order, subsequently, to offer independent shared services related to a given terminology. A distinction should be made, therefore, between the unified meta-model which we shall call UMV2 and the extensions specific to each terminology which we shall call UMV1 x (where x denotes a particular terminology).

II.2. The CISMeF Terminology Database

This system was established around the « Descriptor »which is the central concept of the terminologies (aka « keyword »). Each descriptor is labelled and may be defined, linked to other descriptors (such as Related-Term relation) and involved in a son-father type of hierarchy (BT-NT for Broader Term - Narrower Term). A descriptor may also contain specific attributes and synonyms, abbreviations etc. It was also necessary to work on the terminologies modelling (OWL format) in order to fit it into the global database structure and to standardize the data in a well known and shared format. That is why the RDF (Resource Description Framework) syntax was chosen with the OWL (Ontology Web Language).

II.3. OWL model

The first part consisted of creating a generic OWL model that could include all the terminologies. The Unified Model for Vocabularies has been specified for this requirements. The next stage consisted of creating one model for each terminology. Thus, the original data was collected and its native structure needed to be well understood.

II.4. OWL data files

With the OWL models, the work consisted to develop a parser for each terminology : the input is the original data (or normalized original data) and output is an OWL file. As data could be in different shapes and structures, in some cases additional processes were performed (temporary databases, files etc.).

II.5. Database integration

The final stage is the integration of the OWL files into the CISMeF IS. A generic parser was developed to directly insert each terminology into the database. A special model was designed to represent each terminology in a « CISMeF Terminology Database view »so that the parser can use this model in input to recognize descriptor classes, definitions, synonyms, relations in order to insert it very easily into the database.

II.6. Creation of the HMTP

The HMTP was designed as a graphic interface of a Web Service, entirely dedicated to information retrieval and associations between terms of several terminologies. Thus, the main objective was to dissociate the substance from the form, in particular the interface.

II.7. The HTMP Web Service

This Web Service was the most important part of the task : retrieve information and major schemes to allow the fullest display in the HTMP interface. The HMTP Web Service has been developed to respect Web Services Standards with SOAP (Simple Object Access Protocol) and WSDL (Web Service Description Language) signatures. It presents some methods to search terms by descriptor or by database unique identifier. A specific assessment of SQL queries on the database has been performed to obtain the best performance for an optimized response time. This program queries a special version of the CISMeF IS with extended tables. Another important point of this Web Service is the security management. Axis2 (Apache) is used to deploy Web Services and its module Rampart, which deals with security to authenticate users that want to access the signatures of the PTS Web Service. Finally, the Jena API was used to generate the final SKOS (Simple Knowledge Organization System) file sent by the Web Service as a response. Consequently, this file is well formed and deals with W3C standards.

II.8. The HTMP website

As the HMTP exploits a SKOS file (RDF), the graphic interface that renders the final HTML was build based on JSP (Java ServerPages) files including XSL (eXtensible Stylesheet Language) functions. Additional CSS (Cascading Style Sheets) and JavaScript functions are implemented to offer a better website design. The final HTML rendering is processed by the client navigator. This method is a major positive factor for the web application because it works with a minimum of effort. The website has been developed for Firefox 3.x but also works on Internet Explorer 6 and later, Google Chrome and Safari (it has not been tested yet on Opera). The final output (XHTML) deals with W3C standards. For optimal performance, special AJAX (Asynchronous JavaScript And XML) methods are implemented. Since the whole SKOS file data is not directly displayed on the navigator screen, it is useless to transform the entire document in XHTML with the

Terminologies	27
Concepts	≻ 867,791
Synonyms	≻ 1,837,761
Definitions	223,654
Relations and hierarchies	2,990,365

TAB. 1 – Main figures of the Health Multi-Terminology Portal

XSL. Therefore, with JavaScript methods, it is possible to re-transform specific portions of the SKOS file immediately (e.g. relations, hierarchies, results of research by terminology). This technology is a very powerful way to increase load speed and to reduce the XSLT processor load for the client navigator. It is also very interesting because usually, AJAX utilisation means a direct server request but with the combination of a Web Service, XSLT and AJAX, however this step is not necessary (it also reduces the server load and the transformation speed).

II.9. Hardware, software and standards

The HMTP Web Service responds in SKOS language (from the RDF global model) and deals with Web Services Standards such as WSDL and SOAP. It is written in Java (J2EE on jre 1.6). The HMTP is composed of several servlets that query the different WSDL signatures of the Web Service. The graphic interface is a set of JSP containing XSL functions and templates. Advanced JavaScript methods and CSS are used to finalize characteristics and the client functionalities of the final XHTML webpage. The HMTP has been mainly developed for Firefox 3.x web browsers.

III. RESULTS

Overall, the time spent to build this health multi-terminology portal is approximately seven man-years. Currently, the CISMeF team is using one junior engineer (JG) to integrate new terminologies (e.g. SNOMED CT) and one post-doc (TM) to align each terminology to another (more than 200 alignments will be performed using CISMeF NLP tools). This terminology portal is available at the following URL with a restricted access : http ://pts.chu-rouen.fr/ (click on « Log in » ; id=cismef & password=demo10). Table 1 displays the number of descriptors and relations included in the HMTP. Due to various optimizations, the average response time for one concurrent user takes less than 500 milliseconds. Since January 2010, HMTP is daily used by CISMeF librarians to index in multi-terminology mode.

IV. DISCUSSION-CONCLUSION

The Health Multi-Terminology Portal (HMTP) is used daily by various CISMeF academic partners in different French and European projects. In 2010, the main HMTP target is the health student to learn how to manipulate health terminologies (e.g. about rare disease with Orphanet thesaurus or anatomy with the FMA ontology) and to extract knowledge from it, in particular from hierarchies and relations (e.g. various siblings of a rare disease, symptoms of this rare disease or to obtain all the muscles of the forearm in one click).
A health multi-terminology portal is a valuable tool to help to index and retrieve resources from a quality-controlled health gateway. It can also be very useful for teaching or performing audits in terminology management.

REFERENCES

[1] 3. Darmoni SJ, Leroy JP, Baudic F, Douyère M, Piot J, Thirion B : *CISMeF : a structured Health resource guide. Methods Inf Med 2000*, 39(1) :30-35.

[2] Douyère M, Soualmia L, 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*, 21(4) :253-261.

[3] Pereira S, Névéol A, Kerdelhué G, Serrot E, Joubert M, Darmoni SJ : *Using multi-terminology indexing for the assignment of MeSH descriptors to health resources in a French online catalogue. AMIA Annu Symp Proc 2008 Nov* 6, 586-590.

[4] 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.

[5] Darmoni SJ ; Joubert M ; Dahamna B ; Delahousse J, Fieschi M : *a French Health Multi-terminology Server. AMIA symp 2009*, 808.

[6] Merabti T, Massari P, Joubert M, Sadou E, Lecroq T, Abdoune H, Rodrigues JM, Darmoni SJ : *An Automated Approach to map a French terminology to UMLS. Proc. MEDINFO 2010.*