The Termontography Workbench: a Protégé-based tool for the compilation of multilingual terminological resources

Peter De Baer, Koen Kerremans and Rita Temmerman

Erasmushogeschool Brussel - Centre for Terminology and Communication
(http://cvc.ehb.be/)

In this poster we describe the Termontography Workbench, a tool for the compilation of ontologically structured multilingual terminological resources starting from a multilingual domain-specific corpus.

This tool supports the termontography methodology, a multidisciplinary approach in which theories and methods for a multilingual terminological analysis in accordance with sociocognitive theory [1] are combined with methods and guidelines for ontology engineering. A clear distinction is made between conceptual modeling at a culture-independent level and a culture-specific analysis of units of understanding. Hence, the prototypical structuring of understanding is taken into consideration.

We currently use the workbench for the PoCeHRMOM-project to build a trilingual (Dutch, English and French) terminological resource and application ontology in the field of e-HRM. By providing an online e-HRM platform that includes terminological information on functions, educational background and competencies we wish to improve the matching between job offers and candidate profiles. Furthermore we want to make it easier for SME's to participate in emerging e-HRM processes.

The workbench uses Protégé as an object database. Our model (supporting the TBX-standard) is used to structure the multilingual terminological information and contains the following information:

1. Source: a corpus text document.
2. Term: a language and text string combination.
3. (Term) description: a description of a specific term or a meaningful text phrase (i.e. law).
4. Term relation: a relation between two terms, for example the full form of an abbreviation.
5. Concept: a relevant unit of understanding within the domain of interest.
6. Concept relation: a relation between two concepts, for example a required competency for a certain function.
7. Category: a unit of understanding used to classify sources, terms, (term) descriptions, term relations, concepts, concept relations and categories.

Categories and concepts can be represented by near-synonymous or near-equivalent terms. The categories are structured in a partitive hierarchy while the concepts are structured in a generic hierarchy on the domain. In the workbench both hierarchies are displayed as a tree model.

Our decision to use Protégé as the underlying database was mainly based on the following considerations:

- We wanted the model to be flexible while building the prototype of our workbench without having to hardcode and maintain complex SQL-statements.

- Using the database backend of Protégé we expect the database to be expandable over time.
- Using the server version of Protégé we believe it will be possible to build a multi-user version of the workbench.
- We hope to be able to exchange ontology information with other knowledge engineers using the Protégé tool(s).
- In the future we want to research how ontology inference can support the creation of terminological resources.

[1] R. Temmerman, Towards New Ways of Terminology Description. The sociocognitive approach, John Benjamins, Amsterdam/Philadelphia, 2000.