Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 173 (2015) 264 – 269

Procedia
Social and Behavioral Sciences

32nd International Conference of the Spanish Association of Applied Linguistics (AESLA): Language Industries and Social Change

Methodologies to build ontologies for terminological purposes

María Rosario Bautista-Zambrana[a]*

[a]*Universidad de Málaga, Calle León Tolstoi, s/n, Málaga 29071, Spain*

Abstract

This paper focuses on the construction of ontologies for terminological purposes, and more precisely, on the design criteria and methodologies that can be adopted to this end. Therefore, we will outline some important design criteria and compare some relevant methodologies, with respect to five factors; on the basis of that study, we will propose our own methodology, designed to build ontology-based multilingual terminology resources.

Peer-review under responsibility of Universidad Pablo de Olavide.

Keywords: Terminology; ontologies; methodology; design criteria; multilinguality

1. Introduction

In the last two decades, there has been an increasing interest in terminological knowledge bases (TKB), a term first coined by Meyer et al. (1992) (see Temmerman and Kerremans, 2003). TKBs can be constructed in a number of ways; one of the best-known technologies to do so are ontologies; they have become an important object of study for a number of terminology-oriented research groups, and several periodical conferences have been dealing with this area of expertise for years now: the Terminology & Ontology: Theories and Applications Conference (TOTh), the International Conference on Terminology and Artificial Intelligence (TIA) and the Terminology and Knowledge Engineering Conference (TKE) (see Durán-Muñoz and Bautista-Zambrana, 2013).

Although *Ontology* (with a capital letter) has its origin in Greek philosophy, it first acquired great relevance and widespread use in the field of ontological engineering, subfield of artificial intelligence. In this domain *ontology*

* Corresponding author. Tel.: +34-952133260; fax: +34-952131843.
E-mail address: mrbautista@uma.es

Peer-review under responsibility of Universidad Pablo de Olavide.
doi:10.1016/j.sbspro.2015.02.063

(with a lower case letter) can be defined as "a formal, explicit specification of a shared conceptualization" (Studer et al., 1998, p. 25). The authors explain their definition as follows:

> A 'conceptualisation' refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. 'Explicit' means that the type of concepts used, and the constraints on their use are explicitly defined […]. 'Formal' refers to the fact that the ontology should be machine readable, which excludes natural language. 'Shared' reflects the notion that an ontology captures consensual knowledge, that is, it is not private to some individual, but accepted by a group. (Studer et al., 1998, p. 25)

As we mentioned above, knowledge-based approaches, and especially ontologies have played a significant role in terminology in the past years, when they were adopted from the field of ontological engineering. We will understand *terminology* here as the discipline that studies terms and the concepts underlying them, and also as "the activity of collecting, describing and organizing terms in a resource, such as a term bank or a specialized dictionary" (L'Homme and Bernier-Colborne, 2012, p. 389). (In this second sense, we will also use the designation *terminology work*). The reason for the advantageous relationship between terminology and ontologies has its origin in the conceptual approach of terminology work; specifically, it lies in the fact that ontologies provide a type of formal conceptual structure that can result in more controlled, consistent, logical and systematic terminological resources (Vargas Sierra, 2007; Durán-Muñoz and Bautista-Zambrana, 2013). Along this line of thought, Moreno Ortiz (2008, p. 3) points out the following: the terminologist works with concepts; he draws up and defines in a systematic way the conceptual relations that exist among those concepts, creates taxonomies and specifies which lexical units are used, in the various languages under study, to make reference to those concepts; therefore, the author adds, it seems that a formalized, explicit and standardized conceptual representation system [ontologies] should facilitate part of the terminologist's work.

The article is organized as follows. Section 2 focuses on design criteria and methodologies for constructing ontologies; after giving an account of several important design criteria, six methodologies are compared. Section 3 presents our own methodology for building ontologies for terminological purposes. Finally, the conclusion gives a brief summary of the study and a final remark about the application of our methodology.

2. Methodologies for building ontologies

When creating an ontology, it is important to follow ordered and defined steps, that is, to adopt a methodology. According to de Hoog (1998, p. 2), a *methodology* refers to knowledge about methods, and prescribes how an agent should act (what, how and when) to achieve a certain goal. It is important to distinguish this concept from those of *method* and *technique*: as Greenwood (1973, apud Gómez Pérez et al., 2004, p. 108) explains, the former is a general procedure, while a technique is the specific application of a method and the way in which it has to be executed.

Prior to choosing a methodology to build an ontology, it is necessary to take into account the criteria that will guide its design.

2.1. Design criteria

Several authors have dealt with design criteria, such as Gruber (1995), Gómez Pérez et al. (2004) and Brewster and Wilks (2004). In this respect, Gruber (1995) states that "we need objective criteria that are founded on the purpose of the resulting artifact". We have selected the criteria proposed by these authors that we will take into account for our subsequent work:

- Clarity: Knowledge should be presented in a clear way. This can be achieved by the minimization of the syntactic distance between sibling concepts, the standardization of names and the use of single labels.
- Coherence and consistency: An ontology should "sanction inferences that are consistent with the definitions" (Gruber, 1995, p. 909).
- Extendibility: "An ontology should be designed to anticipate the uses of the shared vocabulary" (Gruber, 1995, p. 909).

- Minimal encoding bias: "The conceptualization should be specified at the knowledge level without depending on a particular symbol-level encoding" (Gruber, 199, p. 909).
- Possibility of multiple inheritance: We must be able to place a concept in multiple positions in the taxonomy.
- Ease of computation: It is important "that the methods chosen do not have great complexity and therefore excessive computational cost" (Brewster and Wilks, 2004, p. 15).

2.2. Comparison of methodologies

We have selected six methodologies whose proposals meet the design criteria mentioned above. Given that ontologies are mainly used in ontological engineering, many of the existing methodologies are geared to the organization and exchange of information in computer systems, as well as in the Semantic Web. Nevertheless, we consider that it is possible to adapt those methodologies to the aims of terminology work. Some of them are, for instance, Uschold and King's (1995), METHONTOLOGY (Gómez Pérez et al., 2004), On-To-Knowledge (Staab et al., 2001) and Noy and McGuinness' (2001). Other methodologies arose from the work by terminology researchers interested in taking advantage of the features of ontologies for creating terminological resources. Two of the most relevant are TERMINAE (Aussenac-Gilles et al., 2008) and Termontography (Temmerman and Kerremans, 2003).

We will compare these methodologies taking into account five criteria that are of interest for terminology work and for terminologists who wish to build ontology-based multilingual terminological resources (for instance, term banks or specialized dictionaries):

- Criterion 1, C1: Ontology based on corpus. Are the ontology elements (concepts, relations, properties, etc.) based on corpus work? (For a more detailed account of ontology components, see Corcho et al., 2005.)
- C2: Intended audience. Who are the intended users of the methodology?
- C3: Level of detail. Does the methodology explicitly state which methods and techniques we should use in order to perform the different activities? If so, we will grade this aspect on a scale from 1 (very little detail) to 5 (thorough description of methods and techniques).
- C4: Associated software application. Is there a program associated to the methodology, which facilitates the different steps to be taken?
- C5: Conceptualization phase. Does the methodology propose to perform a conceptualization activity, as defined by Gómez Pérez et al. (2004)? (According to these authors, *conceptualization* consists in organizing and structuring knowledge, by means of external representations, independent from the knowledge representation paradigms and ontology languages that will be later used to implement the ontology; those representations (for instance, diagrams and tables) must be comprehensible by both domain experts and ontology developers.)

By analyzing the methodologies from this perspective, we expect to determine which methodology is most useful for our objectives, so as to use it as a basis for our own methodology.

Table 1 shows the results of the comparison that we have carried out after examining, for each methodology, the main publication(s) that describe(s) it.

Table 1. Comparison of methodologies.

	C1	C2	C3	C4	C5
Uschold and King	No	Ontology developers	3	No	No
METHONTOLOGY	No	Ontology engineers and researchers	5	WebODE and Protégé	Yes
On-To-Knowledge	No	Ontology developers	4	OntoEdit (now OntoStudio)	Yes
Noy and McGuinness	No	Ontology developers	5	Protégé	Not explicit
TERMINAE	Yes	Knowledge engineers and terminologists	4	Terminae	Yes

Termontography	Yes	Ontology builders, terminographers and lexicographers	3	Termontography Workbench	Not explicit

After reviewing Table 1, we can draw the following conclusions: (We have also taken into account other data from the references cited, but these cannot be explicitly mentioned because of space constraints.)

- The methodologies based on ontological engineering do not contain guidelines about knowledge acquisition from corpus, nor do they lay emphasis on the extraction of terms and relations from linguistic resources. The use of multilingual lexicalizations is not mentioned.
- TERMINAE and Termontography propose the acquisition of knowledge from corpus. Termontography addresses the problem of multilingual knowledge representation and emphasizes the relevance of linguistic, cultural and categorization differences between languages.
- METHONTOLOGY, although not based on corpus work, seems to be particularly advisable because of its detailed instructions and association to ontology editors.
- The methodology proposed by Noy and McGuinness is clear and detailed, but does not offer explicit instructions on conceptualization.
- From the point of view of tool availability, the methodologies aimed at terminology seem to be advantageous, but *Terminae* is only available for projects in English and French, while *Termontography Workbench* is no longer available online.

Based on these results, METHONTOLOGY meets the most criteria, with the exception of corpus-based knowledge extraction. TERMINAE also complies with all the requirements, but its associated tool, *Terminae*, has the drawback of only supporting English and French. None of the methodologies analyzed are completely suitable to the aims that we expressed above, so we propose to create a methodology that combines the best characteristics of METHONTOLOGY, on one side, and of both terminological methodologies, on the other.

3. Proposed methodology

The methodology that we propose takes METHONTOLOGY as a basis and includes methods and techniques based on corpus. The design criteria presented in section 2.1 have also been taken into account.

We propose to perform the following activities: (For a more detailed description of this methodology, see Bautista-Zambrana, 2013.)

- *Specification*. We need to determine why the ontology is being built, and what its intended uses and final users will be.
- *Knowledge acquisition*. This activity relies on corpus work and involves the following tasks:
  a. Compilation of (multilingual) corpus and representativeness study. We propose to compile a comparable textual corpus and determine its representativeness by means of the program *ReCor* (Corpas Pastor and Seghiri, 2010). It is advisable to compile a parallel corpus too.
  b. Term extraction. This can be done with term extractors such as *TermoStat* and *AUTOTERM*.
  c. Detection of linguistic equivalents. We propose to use the bilingual extractor *LexTerm*; in order to find further equivalents, it can be useful to compare contexts manually, using for example a concordancer such as *AntConc*.
  d. Extraction of conceptual relations. We extract them by means of two methods: bottom-up (by analyzing corpus concordances detected by *AntConc*) and top-down (by using *WordNet* and specialized dictionaries).
- *Conceptualization*. This is the activity that relies most on METHONTOLOGY. It comprises the following tasks:
  a. Glossary of terms. It contains all the terms, synonyms and linguistic equivalents that we extracted in the previous phase.
  b. Concept taxonomies. We build them by applying a middle-out method (Uschold and Grüninger, 1996).

c. Concept dictionary. It contains the concepts underlying the terms, and the relations and properties related to them.
d. Definition of ad hoc binary relations in detail. We specify which concepts are linked by each relation.
e. Definition of properties. We specify the name of each property, the concept to which it is applied and the type of value (string, integer, etc.).

- *Implementation*. We propose to implement the conceptualization carried out in the previous phase by means of the program *TopBraid Composer Free Edition*.

5. Conclusions

This paper has given an account of the relevance that ontologies can have for terminology work and has outlined the design criteria that should guide the construction of an ontology for terminological purposes. Next, we have compared six methodologies for building ontologies —four from the domain of ontological engineering and two aimed at terminology work— with respect to five criteria, which we find important in order to build ontology-based multilingual terminological resources. The comparison has shown that METHONTOLOGY and TERMINAE are the most suitable methodologies, but none of them comply perfectly with all the requirements. Therefore, we have proposed a methodology that describes in detail the steps to build an ontology for terminological purposes. The proposed methodology has been applied to build an ontology-based multilingual specialized dictionary on package travel, as reported in Bautista-Zambrana (2013).

Acknowledgements

The research reported in this paper has been carried out in the framework of two R&D projects: 'INTELITERM: Sistema inteligente de gestión terminológica para traductores' (Ref. FFI2012-38881, 2012-2015. MEC) and 'TERMITUR: Diccionario inteligente TERMInológico para el sector TURístico (alemán-inglés-español)' (Ref. HUM2754, 2014-2017. Junta de Andalucía).

References

Aussenac-Gilles, N., Després, S., and Szulman, S. (2008). The TERMINAE Method and Platform for Ontology Engineering from Texts. In *Proceedings of the 2008 Conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge*. Amsterdam: IOS Press.

Bautista-Zambrana, M. R. (2013). *Diseño y elaboración de un glosario ontoterminológico multilingüe para el viaje combinado (español-inglés-alemán)*. Doctoral thesis, University of Málaga (Spain).

Brewster, C., and Wilks, Y. (2004). Ontologies, taxonomies, thesauri: Learning from texts. In M. Deegan (Ed.), *The Keyword Project: Unlocking Content through Computational Linguistics. Proceedings of the Use of Computational Linguistics in the Extraction of Keyword Information from Digital Library Content: Workshop 5-6 February*, pp. 1–36.

Corcho, Ó., Fernández-López, M., Gómez-Pérez, A., and López-Cima, A. (2005). Building legal ontologies with METHONTOLOGY and WebODE. In V. R. Benjamins, P. Casanovas, J. Breuker, and A. Gangemi (Eds.), *Law and the Semantic Web. Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications*, pp. 142–157. Springer-Verlag.

Corpas Pastor, G., and Seghiri Domínguez, M. (2010). Size Matters: A Quantitative Approach to Corpus Representativeness. In R. Rabadán (Ed.), *Lengua, traducción, recepción. En honor de Julio César Santoyo*. León: Universidad.

de Hoog, R. (1998). Methodologies for Building Knowledge-Based Systems: Achievements and Prospects. In J. Liebowitz (Ed.), *Handbook of Applied Expert Systems*. Boca Raton, Florida: CRC Press.

Durán-Muñoz, I., and Bautista-Zambrana, M. R. 2013. Applying Ontologies to Terminology: Advantages and Some Disadvantages. *Hermes-Journal of Language and Communication in Business*, *51*, pp. 65–77.

Gómez-Pérez, A., Fernández-López, M., and Corcho, Ó. (2004). *Ontological Engineering: with examples from the areas of Knowledge Management*. London: Springer-Verlag.

Greenwood, E. (1973). *Metodología de la investigación social*. Buenos Aires, Argentina: Paidós.

Gruber, T. R. (1995). Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal Human-Computer Studies*, *43* (5-6), pp. 907–928.

L'Homme, M-C., and Bernier-Colborne, G. (2012). Terms as labels for concepts, terms as lexical units: A comparative analysis in ontologies and specialized dictionaries. *Applied Ontology*, *7* (4), pp. 387–400.

Meyer, I., Skuce, D., Bowker, L., and Eck, K. (1992). Towards a New Generation of Terminological Resources: An Experiment in Building a Terminological Knowledge Base. In *Proceedings of the 14th International Conference on Computational Linguistics (CO-LING'92)* (pp. 956–960). Nantes, France.

Moreno Ortiz, A. (2008). Ontologías para la Terminología: Por Qué, Cuándo, Cómo. *Tradumàtica*, *6*.

Noy, N. F., and McGuinness, D. L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology* [online]. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05. http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noy-mcguinness.html.

Staab, S., Schnurr, H. P., Studer, R., and Sure, Y. (2001). Knowledge Processes and Ontologies. *IEEE Intelligent Systems*, *16* (1), 26-34.

Studer, R., Benjamins, V. R., and Fensel, D. (1998). Knowledge engineering: Principles and methods. *IEEE Transactions on Data and Knowledge Engineering*, *25* (1-2), 161-197.

Temmerman, R., and Kerremans, K. (2003). Termontography: Ontology Building and the Sociocognitive Approach to Terminology Description. In E. Hajicová, A. Kotešovcová and J. Mírovský (Eds.), *Proceedings of CIL17* (pp. 1–10). Prag: Matfyzpress, MFF UK.

Uschold, M., and Grüninger, M. (1996). Ontologies: Principles, Methods and Applications. *Knowledge Engineering Review*, *11* (2), 93–155.

Uschold, M., and King, M. (1995). Towards a Methodology for Building Ontologies. In D. Skuce (Ed.), *IJCAI'95 Workshop on Basic Ontological Issues in Knowledge Sharing* (pp. 6.1–6.10). Montreal, Canada.

Vargas Sierra, C. (2007). El léxico especializado y las ontologías. In E. Alcaraz Varó, J. Mateo Martínez, and F. Yus Ramos (Eds.), *Las lenguas profesionales y académicas* (pp. 41–52). Barcelona: Ariel.