Enhancing Terminological Knowledge With Upper Level Ontologies

Selja Seppälä
University at Buffalo
Buffalo, NY, USA
seljamar@buffalo.edu

Amanda Hicks
University of Florida
Gainesville, FL, USA
aehicks@ufl.edu

Abstract

In this communication, we advocate the use of upper level ontologies such as the Basic Formal Ontology (BFO) to enhance terminological resources and research. First, we present common issues in ontologized terminological work. Then, we review two projects that illustrate the potential advantages of integrating rigorous formal upper level ontologies. Finally, we discuss possible challenges and conclude with a summary of the benefits that such ontologies can bring to both terminological theory and practice.

1 Introduction

Terminologies encode lexical and background knowledge that experts have about their domain of expertise. These resources can be associated with a more explicit ontology-like representation of the entities in the relevant domain. Such representations may include, for example, non-lexicalized concepts. This extends mere terminologies to more sophisticated knowledge representations. Being language independent, ontologized terminologies have the advantage of integrating multilingual terminologies. When augmented with axioms, they can be used in reasoning systems.

Terminological works, where they refer to ontologies at all, generally use Gruber's definition of an ontology as "an explicit specification of a conceptualization." (Gruber, 1995). Ontologies built on the basis of this definition thus depend on peoples' concepts. As a result, the Gruber approach may lead to several distinct ontological representations of the same domain, whether expressed in the same natural language or in different ones. This definition may also lead to a multiplication of ontological terms expressing categories and relations to represent the same or distinct conceptual systems.

However, a multiplication of ontological metalanguages (categories and relations) tends to create knowledge silos (Smith and Ceusters, 2010). In particular, when these metalanguages are domain-specific. Even within a single domain, using distinct metalanguages can limit interoperability of systems using ontological representations of terminologies. Furthermore, from the terminological research viewpoint, a multiplication of categories and relations hinders the advancement of our understanding of conceptual systems, of the internal structure of terms and definitions, etc. To avoid these limitations, we propose that terminologists developing terminological resources and carrying out research would greatly benefit from using an upper level ontology, such as the Basic Formal Ontology (BFO), to integrate resources and research.

In this communication, we present and discuss existing works integrating upper level ontologies, and underline the main advantages of augmenting terminological knowledge with categories and relations from an upper level ontology such as BFO.

2 Limitations of Ontological Terminologies

As shown in Seppälä (2012), common issues in ontologized terminologies are:

- Lack of rigorously defined categories and relations. The interpretation of the metalanguage is left to our intuitive understanding of the terms used for expressing the used categories and relations.

- *is_a overloading* (Guarino, 1998): the *is_a* relation used for structuring the domain ontology does not distinguish the genuine *is_a* subsumption relation from the *instance_of* relation, and sometimes even from the *part_of* relation.

- Multiplication of domain-specific, sometimes ad hoc, categories and relations.

- When upper level categories are used, limitation to a few top-most categories, which are completed with domain-specific ones (Faber, 2002; Kageura, 2002).

The above limitations result in practical and research-related consequences for terminological works, which can be summarized as follows:

- Confusing and incompatible representations of the same domain.

- Non-interoperable terminologies, which hinders the possibility of sharing and reusing terminological resources.

- Non-generalizable observations of terminological phenomena, which hinders research towards a proper understanding of content-related principles governing term formation, definition composition, and conceptual system organization. This eventually hinders the development of widely (re)usable terminological tools, for example, for creating new terms and writing definitions.

- Non-comparable results of terminological research for lack of a common well-defined domain- and language-independent metalanguage, which hinders the development of a mature integrated science.

These shortcomings can be addressed by adopting well-defined domain- and language-independent upper level categories and relations (ontological metalanguage) of the sort accounted for in formal upper level ontologies.

3 Enhancing Terminologies with Upper Level Ontologies

A formal upper level ontology can be defined as "a representation of the categories of objects and of the relationships within and amongst categories that are to be found in any domain of reality whatsoever." (Spear, 2006)

To illustrate the potential advantages for terminology of using formal upper level ontologies, we describe two projects that integrate such ontologies. There are a few upper level ontologies that can be used by mid-level or domain-specific ontologies to define and relate their categories in a non-ambiguous manner, using logical axioms if needed. The projects described hereafter use, respectively, the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) (Masolo et al., 2001) and the Basic Formal Ontology (BFO) (Arp et al., 2015).

3.1 The KYOTO Project

The KYOTO project aims at representing domain-specific terms in a computer-tractable axiomatized formalism to allow machines to reason over texts in natural language (Vossen et al., 2010). The system developed in this project comprises a platform for multilingual text mining and information extraction that was tested on documents from the environmental domain. The semantics of the terms are defined through the KYOTO ontology which is based on DOLCE. WordNets and specialized vocabularies of different languages are linked to ontology classes on the basis of a mapping of the English WordNet to the KYOTO ontology. "This basic ontology and the mapping to WordNet are used to model the shared and language-neutral concepts and relations in the domain." (Vossen et al., 2010, 4) The system can thus "detect similar data across documents in different languages, even if expressed differently." (Vossen et al., 2010, 2)

In Vossen et al. (2013), the authors extracted statements from texts about the Chesapeake Bay using Kybots, scripts based on ontological and linguistic patterns in annotated text. The results of baseline fact extraction were compared with Kybot extraction and Cterm extraction, both of which utilize the KYOTO ontology. The result was that the baseline and Kybot profiles had high recall, 100% and 91% respectively. The baseline had low precision (18%), whereas the precision of the

Kybot profiles was better, though not optimal, at 31%. In short, leveraging ontological information in domain-specific fact extraction NLP resulted in high recall and improved precision.

3.2 The BFO-Based Ontological Analysis Framework

The second project consists in analyzing the contents of definitions using the categories and relations of BFO (Seppälä, 2012; Seppälä, 2015b). The author puts forward an ontological analysis framework that is domain- and language-independent and that can be used in any kind of terminological conceptual analysis task. The categories and their characteristics are also used as models that serve to predict the contents of definitions. These may be used as templates in tools to help in definition writing.

The results of the pilot study reported in Seppälä (2012; 2015b) show that these BFO-Templates account for about 75% of the contents of definitions of terms from 15 distinct domains. The rest of the definition contents can be described using the BFO categories and relations.

The well-defined BFO vocabulary can thus be used as a metalanguage to describe definition contents, term formation, and the organization of conceptual systems in a way that research findings can be compared and integrated. In practice, BFO-based ontologized terminologies would have the advantage of being interoperable, as it is already the case for the mid-level and domain-specific ontologies (and the corresponding terminologies) that extend BFO, such as the Ontology for Biomedical Investigations (OBI) and the Ontology for Biobanking (OBIB)[1].

4 Possible Obstacles to Use of Upper Level Ontologies

Using upper level ontologies may sometimes prove challenging. Possible issues may be:

- Upper level ontologies evolve and their categories are, at times, still under development. In those cases, it may not be straightforward under which category to place a term.
- Specifications of the upper level ontology may be sparse and lacking, and sometimes too formal (OWL, first order logic) to be easily understood by terminologists.
- An adequate use requires familiarity with the upper level ontology chosen.

A solution to such issues would be to use existing mappings of WordNet to upper level ontologies as aids for integration. A future mapping of WordNet to BFO should facilitate the integration of BFO in terminological projects (Seppälä, 2015a).

5 Conclusion

We saw that ontologized terminologies present a number of shortcomings that can be addressed by integrating a formal upper level ontology. We illustrated the advantages of such an enhancement by reviewing two projects that use such ontologies. To summarize, the main benefits of using a language- and domain-independent upper level ontology are, on the practical side, the possibility to integrate multilingual and multi-domain terminological resources with one another and with information system tools. The latter can thus use the inferences drawn on the basis of the upper level ontology to reason over and manipulate multilingual natural language texts. Using a well-defined formal upper level ontology as a basis for terminological work would make sharing and reuse of terminologies easier: identifying and sharing common terms, constructing new definitions using the same building blocks (information types and logical axioms), etc. Such a framework avoids semantic conflicts and need for mapping.

On the research side, using a well-defined ontological metalanguage allows: carrying out rigorous and comparable conceptual analysis work in terminology; making language- and domain-independent generalizations about term formation, definition content structure, and terminological systems' organization, which can help develop empirically based content standards and writing aid tools; creating comparable research results that contribute to developing a mature integrated terminological science.

[1] For a full list, see `http://ifomis.uni-saarland.de/bfo/users`. For an illustration of interoperability and its advantages, see the presentation on *The OBIB Ontology for Biobanking*, by Chris Stoeckert, Jie Zheng, and Mathias Brochhausen `http://ncorwiki.buffalo.edu/index.php/CTS_Ontology_Workshop_2015`.

Moreover, a metalanguage using the categories and relations of an upper level ontology for describing terminological data (for example, terms', definitions', and conceptual systems' structure) can fruitfully complement any terminological resource whether or not already ontologized. Cimiano et al. (2011) propose, for example, a model to formally link lexicons (with relevant linguistic descriptions) to ontologies.

Using more specifically a BFO-based metalanguage would further enhance our understanding of the relationship between the lexical, linguistic, conceptual, and ontological levels of terminologies. Indeed, BFO is a realist ontology that represents the things that exist in the world and the relations between them, independently of our conceptualizations thereof. A BFO-based metalanguage may thus provide an additional level of understanding to existing descriptive frameworks.

We therefore encourage terminologists to fully embrace the best ontological practices to enhance their research and resources.

Acknowledgments

This work was supported in part by the Swiss National Science Foundation (SNSF) and by the NIH/NCATS Clinical and Translational Science Awards to the University of Florida UL1 TR000064. The content is solely the responsibility of the authors and does not necessarily represent the official views of the SNSF, the National Institutes of Health, or the NCTE. Thanks also to Aurélie Picton and Barry Smith.

References

Robert Arp, Barry Smith, and Andrew D. Spear. 2015. *Building Ontologies with Basic Formal Ontology*. MIT Press, Cambridge, MA.

Philipp Cimiano, Paul Buitelaar, John McCrae, and Michael Sintek. 2011. LexInfo: A declarative model for the lexicon-ontology interface. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(1):29–51.

Pamela Faber. 2002. Terminographic definition and concept representation. In Johann Haller Belinda Maia and Margherita Ulyrich, editors, *Training the Language Services Provider for the New Millennium*, pages 1–14 [343–354]. Universidade do Porto, Porto.

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

Nicola Guarino. 1998. Some ontological principles for designing upper level lexical resources. In *Proceedings of First International Conference on Language Resources and Evaluation. ELRA-European Language Resources Association, Granada, Spain*, pages 527–534. Citeseer.

Kyo Kageura. 2002. *The Dynamics of Terminology: A Descriptive Theory of Term Formation and Terminological Growth*. Terminology and Lexicography Research and Practice 5. John Benjamins, Amsterdam.

Claudio Masolo, Stefano Borgo, Aldo Gangemi, Nicola Guarino, and Alessandro Oltramari. 2001. Wonderweb deliverable D18 ontology library (final). *ICT Project*.

Selja Seppälä. 2012. *Contraintes sur la sélection des informations dans les définitions terminographiques: vers des modèles relationnels génériques pertinents*. Ph.D. thesis, Département de traitement informatique multilingue (TIM), Faculté de traduction et d'interprétation, Université de Genève.

Selja Seppälä. 2015a. Mapping WordNet to the Basic Formal Ontology using the KYOTO ontology. In *Proceedings of ICBO 2015*.

Selja Seppälä. 2015b. An ontological framework for modeling the contents of definitions. *Terminology*, 21(1):23–50.

Barry Smith and Werner Ceusters. 2010. Ontological Realism: A Methodology for Coordinated Evolution of Scientific Ontologies. *Applied Ontology*, 5:139–188.

Andrew D. Spear, 2006. *Ontology for the Twenty First Century: An Introduction with Recommendations*. Institute for Formal Ontology and Medical Information Science, Saarbrücken, Germany.

Piek Vossen, German Rigau, Eneko Agirre, Aitor Soroa, Monica Monachini, and Roberto Bartolini. 2010. KYOTO: an open platform for mining facts. In *Proceedings of the 6th Workshop on Ontologies and Lexical Resources*, pages 1–10.

Piek Vossen, Eneko Agirre, German Rigau, and Aitor Soroa. 2013. Kyoto: A knowledge-rich approach to the interoperable mining of events from text. In *New Trends of Research in Ontologies and Lexical Resources*, pages 65–90. Springer.