Creating a National Content and Service Infrastructure for the Finnish Semantic Web

Eero Hyvönen and **Arttu Valo** and **Katri Seppälä** and **Tomi Kauppinen** and
Ville Komulainen and **Tuukka Ruotsalo** and **Mirva Salminen** and **Anu Ylisalmi**
University of Helsinki, Department of Computer Science and
Helsinki University of Technology, Media Technology and HIIT
P.O. Box 5500, Otaniementie 17, FI-02015 TKK
FirstName.LastName@tkk.fi

Abstract

We present a national ontology development and service framework being developed in Finland in 2003-2007. Our goal is to initiate and support collaborative ontology development processes of various expert groups now developing keyword thesauri. The framework is based on a set of related core ontologies, most notably on a national upper ontology based on the commonly used Finnish General Thesaurus YSA. The framework implements three ontology services by a web-based system ONKI. Firstly, ONKI supports distributed collaborative development and versioning of interdependent ontologies. Secondly, external cataloging and indexing systems can use ONKI as a web service for ontology-based annotations. Thirdly, information retrieval systems can use ONKI for disambiguating keyword meanings for concept-based search on the Semantic Web.

1 Introduction

Semantic interoperability on the semantic web is based on ontologies [3]. In focused application and domain areas it may be possible to agree upon common ontological concepts, but on larger cross-domain applications, this usually becomes more difficult. Different domains may need different ontological representations even for the same objects of discourse, and different parties tend to have different philosophical opinions on how to model the world. As a result, there is the threat that the Semantic Web will become a set of isolated, mutually incompatible web islands.

There are various complementary approaches for making semantic web ontologies interoperable. First, ontology mapping and alignment [4] can be used for mapping concepts with each other. However, this is known to be difficult. Second, ontologies can share and be based on common foundational logical principles, like in DOLCE [5,17]. This easily leads to complicated logical systems (e.g., modal logics may be needed) that may not scale up to real word practical usage. Third, horizontal top ontologies, such as the IEEE SUMO[1]

can be created for bridging the concepts between vertical domain ontologies. Fourth, ontology engineering support systems for creating ontologies in the first place as interoperable as possible can be created.

In this paper, we present a national ontology development and application framework and project "Finnish National Ontologies on the Semantic Web" (FinnONTO)[2]. This framework is mainly based on the third and fourth approaches above. The goals of FinnONTO are the following:

1. **From thesauri to ontologies**. The general idea is to move ahead from developing national thesauri [6] to developing ontologies.

2. **Collaborative ontology development**. Develop a national framework for distributed collaborative ontology development.

3. **Core ontologies**. Develop initial versions of a set of central national core ontologies in order to initiate ontology development processes. The most central ontology is the top ontology YSO based on the Finnish General Thesaurus YSA [18]. Resources in YSO will be used and shared by the other interdependent vertical domain ontologies.

4. **Usage as public web services**. Enable ontology usage, especially in indexing and information retrieval, through public web-services. In the following these goals are discussed in more detail.

2 From Thesauri to Ontologies

FinnONTO encourages organizations to start developing ontologies instead of thesauri. The reasons for this are obvious: ontologies can be interpreted not only by humans but also by the machine [7], can hence be used for more accurate indexing and information retrieval, and provide a basis for making information systems semantically interoperable. Even with little extra work, e.g. by just systematically organizing concepts along subclass-of and part-of hierarchies, substantial benefits can be obtained, as demonstrated e.g. in [2].

---

[1]http://suo.ieee.org

[2]http://www.cs.helsinki.fi/group/seco/ontologies

3 Towards a Shared Upper Ontology

We share the vision of the IEEE SUMO project[3]: a shared upper ontology is needed for enhancing semantic interoperability between various domain ontologies. In Finland YSA is widely used for content indexing in libraries, museums, and archives of various kinds both in public and in the industry. YSA contains some general 23,000 terms and is organized as a typical thesaurus [6] including some semantic relations, such as Narrower Term (NT) and Broader Term (BT). In addition, the terms are divided into 61 domain groups, such as Physics, History etc. Since the terms of YSA are used in various vertical domain ontologies, YSA can be considered as a kind of terminological glue between many other Finnish thesauri.

In our work, YSA is being developed into an ontology called YSO. Based on it, a national upper ontology conforming to indexing practices of various content providers could be created. The work concentrates on enriching the semantic information of YSA and for providing better disambiguation of the concepts/terms. In Finland, significant amounts of data have already been annotated using YSA terms and could potentially be exploited on the Semantic Web.

The most central (sub)ontology in YSO will be "events" that roughly corresponds to verb-like concepts such as "buying" or "learning". The general idea is that the other ontological re-sources are extensively used in thematic roles of events. The thematic roles include, e.g., the "agent", "instrument", and "location" of an event. This is an approach widely used in ontological knowledge representation [10].

4 Ontology Library Web Services

In FinnONTO a web-based ontology library system called ONKI is being developed [8] for collaborative ontology development. Maintaining large cross-disciplinary vocabularies such as YSA is difficult for any single editorial group. We envision that ONKI will enable and support distribution of this work to communities, each of which working in its own domain of expertise and with a good motivation for developing its own vocabulary.

ONKI provides services for three user groups:

1. For **ontology developers**, ONKI provides the collaborative ontology development and versioning environment (cf. fig. 1).

2. For a **content indexer**, ONKI provides a web-based browser for finding desired concepts and for transporting the corresponding URI from the ONKI server into an external application.

3. For an **information searcher**, ONKI browser can be used for finding and disambiguating keyword meanings, and for transporting the corresponding URIs into search engines and other applications.

For example, by typing in "bank" the browser finds the different meanings (e.g., bank as a financial institution and river bank) of the word and shows them to the user. After this the right intended meaning can be selected by clicking on it. As a side effect, the corresponding URI is read into the application and can be used for searching. Using such concept-based search is feasible in applications such as [2] supporting ontology-based information retrieval.

5 Discussion

The work presented in this paper is being accomplished during 2003-2007. At the moment, only some parts of the collaborative ontology development framework have been implemented [8] and YSO development is still underway. First version of the ONKI Browser for indexing and and information search has been created and is used internally in the project.

The ontologies and results of FinnONTO will be published open source.

Acknowledgments

References

1. G. Antoniou and F. van Harmelen, A Semantic Web Primer. The MIT Press, 2004.

2. E. Hyvönen, M. Junnila, S. Kettula, E. Mäkelä, S. Saarela, M. Salminen, A. Syreeni, A. Valo, and K. Viljanen: Museum-Finland – Finnish museums on the Semantic Web. User's perspective. Selected papers, Proceedings of Museums and the Web 2004 (MW2004), March 29 - April 3, 2004, Virginia, USA.

3. D. Fensel, Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce. Springer-Verlag, 2004.

4. A. Hameed, A. Preese, D. Sleeman, Ontology reconciliation. In: [5]. 5. S. Staab, R. Studer (eds), Handbook on Ontologies. Springer-Verlag, 2004.

6. D. Foskett, Thesaurus. In: Encyclopaedia of Library and Information Science, vol. 30, Marcel Decker, New York, 1980.

7. M. van Assem, M. R. Menken, G. Schreiber, J. Wielemaker, and B. Wielinga, A Method for Converting Thesauri to RDF/OWL. Proceedings of ISWC-2004, Hiroshima, Japan. Springer-Verlag, 2004.

8. V. Komulainen, A. Valo, and E. Hyvönen, A collaborative ontology development and service framework ONKI. Proceedings of Int. Conf. on Dublin Core and Metadata Application (DC-2005), short papers, Madrid, 2005.

9. L. Stojanovic, Methods and Tools for Ontology Evolution. PhD thesis, Universität Karlsruhe, 2004.

10. J. Sowa. Knowledge Representation. Logical, Philosophical, and Computational Foundations. Brooks/Cole, 2000.

---

[3]http://suo.ieee.org