Finnish National Ontologies for the Semantic Web - Towards a Content and Service Infrastructure

Eero Hyvönen, Arttu Valo, Ville Komulainen, Katri Seppälä,
Tomi Kauppinen, Tuukka Ruotsalo, Mirva Salminen, and Anu Ylisalmi
Helsinki University of Technology and University of Helsinki
Laboratory of Media Technology
P.O. Box 5500, FI-02015 TKK, FINLAND
FirstName.LastName@tkk.fi
http://www.cs.helsinki.fi/group/seco/ontologies

Abstract:
We present a national ontology development and service framework being developed in Finland in 2003-2007. 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 maintained by the National Library of Finland. 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.

Keywords:
Ontologies, semantic web, collaboration, web services

1. Introduction

Metadata standards typically define the properties to be used for content descriptions. For example, Dublin Core[1] lists 15 elements (i.e., properties) such as Title, Creator and Subject. Content interoperability across different application domains is obtained by using commonly agreed elements.

Metadata standards are essential for the Semantic Web [1], too, but more powerful methods for describing semantic content and for obtaining semantic interoperability are needed. Here a central problem is the standardization of the *values* of the standard properties/elements. For semantic interoperability on the web, large shared reference ontologies are needed. For example, in the MuseumFinland[2] [2] system, the values of the Creator and other properties of a collection artifact are taken from a set of seven ontologies. They contain some 10,000 resources that define the meaning of individual persons, organizations, artifact types, locations, actions and other objects. Their meaning is shared between the different museums providing the collection metadata content. Other reference ontologies [3] have been proposed for different purposes. For example, in the Open Directory Project[4] the reference ontology contains over 590,000 categories.

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

[1] http://dublincore.org

[2] http://museosuomi.cs.helsinki.fi

[3] Cf. e.g., CYC http://www.cyc.com, TAP http://tap.stanford.edu/, and SUMO http://suo.ieee.org.

[4] http://www.dmoz.org

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] [5]. 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[6] 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). This framework is mainly on the third and fourth approaches above. The goals of FinnONTO are the following:

1. **From thesauri to ontologies**. The general idea is 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 general Finnish keyword thesaurus YSA[7]. 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 [7]. The reasons for this are obvious: ontologies can be interpreted not only by humans but also by the machine [7], and hence be used for more accurate indexing and information retrieval, and for making information systems semantically interoperable. Even with little extra work, e.g. by just systematically organizing concepts along hyponymies and partonymies, substantial benefits can be obtained, as demonstrated e.g. in [2].

3. Supporting Collaborative Ontology Development

Thesauri are widely used for harmonizing content indexing. Different fields have thesauri of their own. The thesauri are typically developed by domain specific expert groups without much systematic collaboration with other fields. When using such thesauri in cross-domain environments, such as the web, semantic problems arise, e.g., due to ambiguity of literal expressions. For example, in the finance domain the term "bank" has an obvious meaning as an institution, but when considering the nature domain, it has another meaning.

In semantic web ontologies the ambiguity problem is solved by dealing with unambiguous resources identified by URIs instead of literal words. However, support is needed for sharing the URIs across domains. If one needs to define the notion of "river bank", (s)he should be aware of not to mix this concept with "money bank". On the other hand, if one is defining the notion of "blood bank", (s)he could use the more general notion of "bank" and modify it, thus sharing this common notion with other kind of banks considered in other ontologies.

In FinnONTO a web-based ontology library system called ONKI is being developed [8] for collaborative ontology development. The ONKI architecture and publishing process is depicted in figure 1. Three core components of the system are the development repository (ONKI GORepository) for ontologies being edited, the public ontology library containing the set of published interrelated ontologies (ONKI Library), and the browsing service (ONKI Browser) for using the ontologies.

[5] http://www.loa-cnr.it/DOLCE.html

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

[7] http://vesa.lib.helsinki.fi

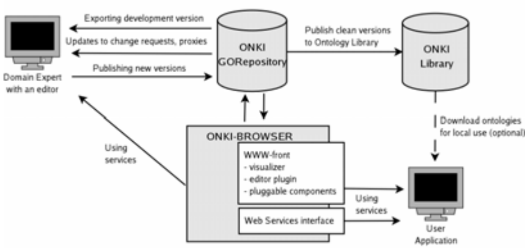

Figure 1. ONKI system

ONKI separates the development process into two major parts: the development loop (cf. the arrows between Domain Expert and ONKI GORepository) and the publishing push (cf. the bottom arrow from Domain Expert to ONKI GORepository and the arrow from there to ONKI Library).

The development loop is based on exporting an ontology with versioning metadata and then occasionally polling the development repository to check, whether there have been affecting changes in other ontologies. Pull is used to download the affecting changes made in related ontologies because it has been identified as a better mechanism for keeping distributed ontology copies and development synchronized [9, p. 152].

4. Towards an Upper Ontology

In FinnONTO, we share the vision of the IEEE SUMO project: a shared upper ontology is needed for enhancing semantic interoperability between various domain ontologies.

In Finland the Finnish General Thesaurus YSA is widely used for content indexing in libraries, museums, and archives of various kinds both in public and in the industry. It contains some general 23,000 terms and is organized as a typical thesaurus [6] including some semantic relations, such as Narrower term (NT) ja 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. In this way 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 been already annotated using YSA terms and could potentially be exploited on the semantic web.

The ontologization process is being performed in the following major steps:

1. Transformation into RDF(S). The thesaurus stored originally in a database in the MARC-format was transformed into 61 RDF(S)[8]. More specifically, projects for the Protégé-2000 editor[9], one for each domain group, were created by transformation scripts. The division of terms/concepts into groups makes it possible for several people to work with the subontologies in parallel.

2. Hyponymy construction. An initial hyponymy for each group was developed with a small set of top categories, such as “locations”, “processes”, “times”, and “qualities”. At this phase, the LT/NT-relations of the thesaurus terms were transformed mechanically into ontological rdfs:subClassOf and rdf:type relations, and transitivity errors were aligned. For example, in YSA the term “pine oil” is a narrower term (NT) of “oil”, that is a narrower term of “diggings”. If the NT relations is changed into the ontological rdfs:subClassOf relation, then the ontology would mean that “pine oil” is a subclass of “diggings”, which is not true. The solution here is to divide the meaning of the term “oil” into two concepts, “mineral oil” and “natural oil”, and to build the hyponymy accordingly.

3. Ontology consolidation. Next, resources belonging to the different top categories are consolidated from the 61 groups, and the hyponymies in the global sense are checked again. At this point, YSO is essentially divided into a set of subontologies corresponding to the roots of the subontologies. At the moment, the project has entered this phase and we are finalizing the subontologies and editing the hyponymies by hand.

4. Ontology enrichment. After phase (3), additional selected semantic relations, such as meronymy and semantic roles, can be added in the ontologies by hand. Since properties are used as a criterion for creating hyponymies, partonomies, and troponymies, this phase partly overlaps (3).

---

[8] http://www.w3.org/RDF/

[9] http://protege.stanford.edu

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 resources 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].

5. Ontology Library Web Services

The YSO ontologies and the set of related ontologies will be published by a public Ontology Library Service ONKI. It 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 of the word and shows them to the user. After this the right intended meaning can be selected the 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.

6. 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 has been created and is used internally in the project.

Acknowledgments

Thanks to Markus Holi and Miikka Junnila who participated in the project earlier, and Juha Hakala for support in obtaining the YSA thesaurus in electronic form. The FinnONTO-project is funded by the National Technology Agency Tekes and a consortium of 14 public and private organizations, including the National Library of Finland maintaining the YSA thesaurus.

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: MuseumFinland—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. Encyclopaedia of Library and Information Science, Volume 30. Marcel Decker, New York, 1980.
7. M. van Assem and M. R. Menken and G. Schreiber and 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, E. Hyvönen, A collaborative ontology development and service framework ONKI. Paper, Helsinki University of Technology, Laboratory for Media Technology, 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.