Semantic Web Applications in the Public Sector in Finland

– Building the Basis for a National Semantic Web Infrastructure

Eero Hyvönen
Helsinki University of Technology (TKK) and University of Helsinki
Laboratory of Media Technology
Semantic Computing Research Group (SeCo)
P.O. Box 5500
FIN-02015 TKK, FINLAND
eero.hyvonen@tkk.fi
http://www.seco.tkk.fi/

1. Towards a National Semantic Web Infrastructure

This article gives an overview of the Finnish National Semantic Web Ontology project (FinnONTO), 2003-2007 (http://www.seco.tkk.fi/projects/finnonto/). Its ambitious goal is to lay a foundation for a national metadata, ontology, and ontology service framework in Finland, and demonstrate its usefulness in practical applications (Hyvönen et al., 2005b). In our vision, a conceptual semantic infrastructure is needed for semantic web services in the same way as roads are needed for traffic and transportation, power plants and electrical networks are needed for energy supply, or GSM standards and networks are needed for mobile phones and wireless communication. A solid, commonly agreed open infrastructure would make it much easier and cheaper for public organizations and companies to create interoperable, intelligent services on the coming semantic web (http://www.w3.org/2001/sw/). The infrastructure should be open source and its central components be maintained by the public sector order to guarantee wide usage and interoperability across different application domains and users.

The consortium behind the project now includes 30 public organizations and companies funding the research, and 8 new members plan to join the team in autumn 2006. This consortium represents a wide area of functions of the society including museums, libraries, health organizations, government, media, and education. Public organizations, companies, and universities are participating in the project. The research is directed and is mostly carried our by the Semantic Computing Research Group (SeCo) (http://www.seco.tkk.fi/) at the Helsinki University of Technology (TKK) and the University of Helsinki. Also the University of Tampere is contributing to the work.

The goals of the project are the following:

1. **Metadata standards**. Nationally adapted standards for representing metadata in various applications fields are needed and are being created.
2. **Core ontology library**. We develop initial versions of a set of central national core ontologies in order to initiate ontology development processes in Finland. The idea is that after the research period, the participating organizations could continue developing the ontologies for machine and human usage instead of traditional thesauri. The most central ontology is the top ontology YSO based on the general Finnish keyword thesaurus YSA (some 23,000 terms) (http://vesa.lib.helsinki.fi), maintained by the National Library of Finland. Resources in YSO will be used and shared by a set of other, mutually interdependent vertical domain ontologies.
3. **Public ontology services**. An ontology library and web service framework ONKI is being developed (Komulainen et al., 2005) to enable ontology usage in ontology development, indexing, and information retrieval, through public web services.
4. **Open source tools.** An tool developed within project is OntoViews (Mäkelä et al., 2004), a framework for creating semantic search and browsing services based on the multi-facet search paradigm. A key bottleneck of the proliferation of the semantic web is production of metadata. For

this purpose, a number of semiautomatic content annotation tools are being developed.

5. **Pilot applications.** The above framework is being evaluated by implementing a number of practical applications in the domains of eCulture, eHealth, eGovernment, eLearning, and eCommerce.

In goals 1-3, semantic web recommendations and best practices of using them are being developed as national adaptations of international similar efforts. We envision that provision of these results as usable public web services is needed to share contents, enforce common practices, and to support organizations in utilizing the new technologies. The project also eats its own "dog food" (goal 5) by creating a number semantic portals and applications using the technology developed – these systems provide a public test bed for evaluating the usefulness of the infrastructure developed. For this purpose, a number of general purpose tools (goal 4) are being developed for content creation, semantic search, and browsing.
In the following these goals and results obtained thus far are discussed in more detail.

2. Metadata Standards

Metadata standards typically tell what properties to use for content descriptions. For example, Dublin Core (http://dublincore.org) lists 15 elements such as Title, Creator and Subject. Content interoperability across different application domains is obtained by using commonly agreed elements.
In FinnONTO, metadata standards are being developed and adapted in several application fields. In the eCulture domain, a metadata scheme for representing museum artifact collection metadata was developed and is in use in the semantic portal "MuseumFinland – Finnish Museums on the Semantic Web" (http://www.museosuomi.fi) (Hyvönen et al., 2005a). The scheme is being developed further in a follow-up system "CultureSampo – Finnish Culture on the Semantic" that addresses the problem of semantic interoperability of different kinds of cultural contents. In geoinformatics, an official national metadata recommendation for spatial information has just been finished (JHS 158, 2006). In eHealth, metadata for health promotion and services is in focus, and in eLearning, a national scheme based on the Learning Object Metadata (LOM) standard is being developed for a general semantic learning portal.
Metadata standards are essential for the semantic web but not enough for semantic interoperability. Here the problem is standardization of the *values* of the commonly agreed elements. For semantic interoperability on the web, large shared reference ontologies are needed. For example, in the MuseumFinland 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 etc. Their meaning is shared between the different museums providing the collection metadata content. This idea is similar to the Open Directory Project (http://www.dmoz.org), where the reference ontology contains over 590,000 categories.
Semantic interoperability on the semantic web is based on ontologies (Fensel, 2004; Staab, Studer, 2004). In focused domains and applications it may be possible to agree upon common ontological concepts, but on a larger cross-domain setting, 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 and needs on how to model that world. As a result, there is the danger that the global semantic web will not emerge but rather a set of isolated, mutually incompatible semantic web islands.
There are various complementary approaches for making semantic web ontologies interoperable. First, ontology mapping and alignment (Hameed et al., 2004) can be used for mapping concepts with each other. Second, ontologies can share and be based on common foundational logical principles, like in DOLCE (http://www.loa-cnr.it/DOLCE.html). This easily leads to complicated logical systems (e.g., modal logics may be needed) that may not scale up either epistemically or computationally to real word situations and practical usage. Third, horizontal top ontologies, such as the IEEE SUMO (http://suo.ieee.org/) 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.

3. Core Ontology Library

Why Thesauri Are Not Enough But Ontologies Are Needed?

FinnONTO encourages organizations to start transforming thesauri into ontologies (van Assem et al., 2004). The problem with thesauri, such as YSA, is that its semantic relations have constructed mainly to

help the indexer in her work. Understanding the relations needs human knowledge. The computer could make use of the semantic structure in many application areas, such as semantic search, information retrieval, semantic linking of contents, automatic indexing, and in making contents semantically interoperable. However, this is difficult, because the computer does not have the human knowledge of an indexer. Ontologies define the meaning the indexing terms and concepts accurately enough for the machine to use, and can hence be used for more accurate indexing. Even with little extra work, e.g. by just systematically organizing concepts along class hierarchies and partonymies, substantial benefits can be obtained, as demonstrated e.g. in (Hyvönen et al., 2005a).

For example, consider the following example translated from the YSA thesaurus, where BT indicates the "broader term" relation used in thesauri:

```
Halley's Comet BT comet
comet BT solar system
```

We can easily understand its meaning but the machine is confused: Is Halley's Comet an individual or a class of them, such as comet? Can there be many Halley's comets or only one? Is a comet a kind of solar system or a part of a solar system? Is it a part as a concept or are all individual comets a part of some solar system? Do comets have the properties of solar systems, e.g. own planets, based on the BT relation. Should a search for solar systems retrieve comets although comets are not solar systems?

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 or musical instrument domains, there are other meanings.

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 and users. If one needs to define the notion of "river bank", one 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", it is possible to use the more general notion of "bank" and modify it, thus sharing this common notion with other kind of banks considered in other ontologies.

A Method for Transforming Thesauri into Ontologies

In our work, YSA is being developed into the national top ontology YSO. Since YSA is widely used in Finnish organizations, YSO would be an important step in solving semantic indexing and interoperability problems in Finland. The thesaurus-to-ontology transformation of YSO is being performed by enriching the semantic of structures of YSA in the following way:

1. Disambiguate individuals from classes. In our example above, Halley's comet would be an individual of the type celestial body. It is not a solar system but a part of a solar system.
2. Disambiguate major concept meanings. For example, in YSA there is the indexing term "child". However, this term has several meanings such as "a certain period of human life" or "a family relation". For example, I am not a child anymore but still a child of my mother. The computer cannot understand this unless the meanings are separated and represented as two entries (URIs) in completely different parts of the ontology. However, also the general concept of child encompassing both meanings can be useful for indexing purposes when both meanings are intended at the same time.
3. Refine and disambiguate the major meanings or BT, NT (narrower term), and RT (related term) relations. For example, the BT relation between Halley's comet and solar system is an instances-class relation while the BT relation between comet and solar system is a part-of relation. We consider, in the first version of YSO, only simple type, subclass, and part-of relations.
4. Check transitivity of semantic relation chains. For example, if an individual is an instance of a class, then it is a member of all its superclasses. In our example, Halleys' comet is a comet but is not a solar system although there is a BT relation chain between the two concepts.
5. Reorganize and a complete the structure into a simple taxonomic ontology. All concepts in the ontology should be connected into a class structure in order to facilitate inference about their relations, and in order to provide the end-user with useful classifications.

In FinnONTO, we share the vision of the IEEE SUO Working Group (http://suo.ieee.org): a shared top ontology is needed for enhancing semantic interoperability between various domain ontologies. In Finland the YSA thesaurus 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. 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 semantic terminological "glue" between many other Finnish thesauri. Once the structure of the top ontology is defined, the same choices can be reused in many cases in the vertical ontologies that typically share lots of concepts with the top ontology. For example, when we created a cultural ontology MAO based on the Finnish Cultural thesaurus MASA (Leskinen, 1997) of some 6,000 terms, about 2,000 terms could immediate be mapped on YSA. We now work on the Argiforest thesaurus (7000 terms) where thousands of terms originate from YSA, too (http://www-db.helsinki.fi/agri/agrisanasto/Welcome_eng.html).
Our core ontology library consists of a set of mutually interdependent ontologies. Support is needed in order manage the interdependencies and for using relevant subsets of ontologies in indexing and information retrieval.

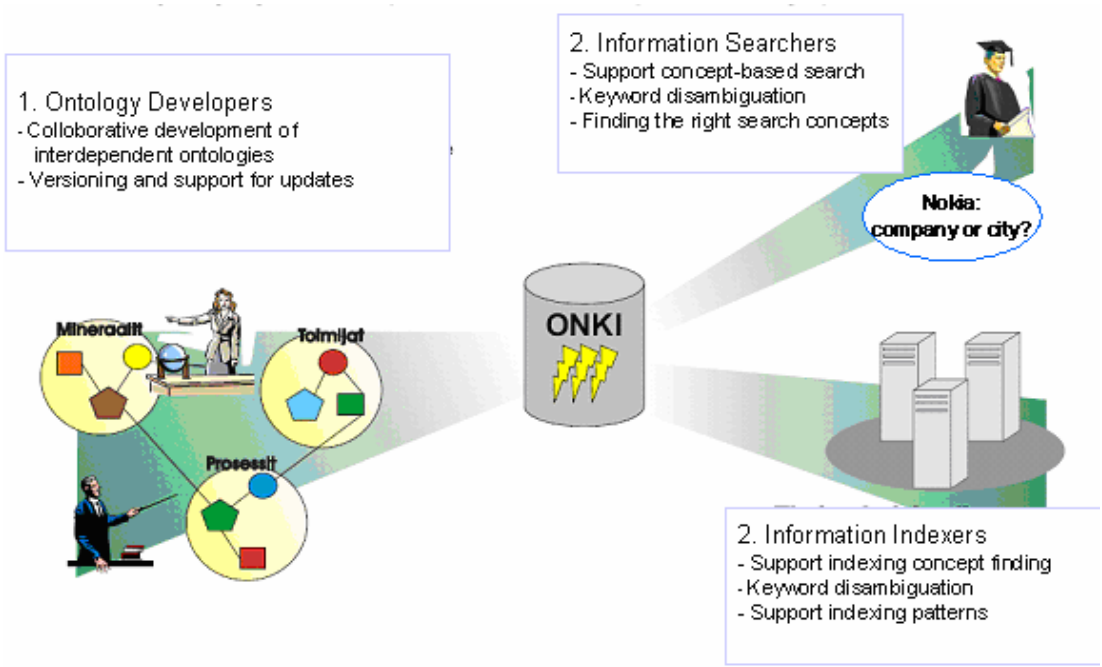

Figure 1. ONKI ontology library provides services to three user groups.

4. Public Ontology Services

The YSO ontologies and the set of related ontologies will be published by a public Ontology Library Service ONKI (Komulainen et al., 2005). It provides services for three user groups (figure 1):

1. For ontology developers, ONKI provides the collaborative ontology development and versioning environment (Valo et al., 2005).
2. For content indexers, ONKI provides a web based browser for finding desired concepts and for transporting the corresponding URI from the ONKI server into the external application.
3. For 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 MuseumFinland supporting ontology-based

information retrieval.

A demonstration illustrating ONKI's indexing services is available at: http://www.seco.tkk.fi/applications/onki/.

4. Open Source Tools

The project has developed the OntoViews tool (Makelä et al., 2004) for creating semantic portals. In our work we have generalized the multi-facet search paradigm for using semantic web ontologies, reasoning, and standards. The first application demonstrating the usability of the scheme was the MuseumFinland portal that was awarded with the international Semantic Web Challenge Award (2. prize) in 2004. The software, as used in this portal, is available at http://www.seco.tkk.fi/projects/semweb/dist.php.

Since then, the tool has been extended and applied in various other demonstrational portals (Mäkelä et al., 2005a), such as Orava (http://www.museosuomi.fi/orava/), a semantic portal containing over 2200 video clips and learning objects provided by the National Broadcasting Company YLE. Orava is a semantic web version of the Klaffi-portal (http://www.yle.fi/klaffi/) of YLE.

The project also develops tools for semiautomatic content annotation, such as Terminator used for term extraction, and Annomobile for matching keywords with ontology URIs (Hyvönen et al., 2005a). At the moment, natural language processing techniques are being developed for creating a tool for annotating text documents.

5. Pilot Applications

The technology developed has and is being applied to case studies in several application domains in order to test its feasibility:

1. **eCulture. "**CultureSampo – Finnish Culture on the Semantic Web" is the next generation of the MuseumFinland portal. The idea of the system is investigate how cultural contents of different kinds could be made semantically interoperable and be published by a shared publication channel on the semantic web. Content types under study include photographs, fine art, videos, folk poetry, documents, such as manuscripts and biographies, web pages, and cultural process descriptions, such as handcraft and farming.
2. **eHealth.** Tervesuomi.fi is the pilot version of a coming national health promotion portal that is being developed in a larger project managed by the National Public Health Institute (KTL). The contents of the portal will be created in a distributed way by a variety of Finnish health organizations facilitated by the tools provide by FinnONTO. The OntoViews framework is used for implementing the portal. A research topic here is to study how content from other content repositories and portals, such as the governmental citizen's portal Suomi.fi, could be integrated automatically with Tervesuomi.fi. Reusing contents would eliminate unnecessary duplication of content work in various governmental organizations, and enrich the services from the end-user's viewpoint. We also do research on uncertainty in ontologies (Holi, Hyvönen, 2006a) and fuzzy ontology mapping in order map end-user's vocabularies with complicated medical ontologies (Holi, Hyvönen, 2006b).
3. **eGovernment.** Suomi.fi (http://www.suomi.fi) is the official citizen's information portal provided and maintained by the Ministry of Finance and a number of public organizations in Finland. We created a demonstration of a semantic version Suomi.fi (Sidoroff, Hyvönen, 2005). The main idea here was to show, that by using shared ontologies, metadata of the web contents and services of different organizations could be integrated automatically into a seamless repository. Different orthogonal views to the content can be provided for the end-user, and mutually related data aggregated along the views and categories in them. For example, the end-user may search the contents based on a classification view of major events in human life, such as "Baby is born". Information related to this event can automatically be gathered from a variety of heterogeneous organizations of heath care, social support, legal administration, church etc.
4. **eLearning.** Opintie is an extension of the current Orava system created for the video and learning object materials of the national Finnish Broadcasring Company YLE. Our goal is to create a demonstration of an open publication channel for learning materials. As in Tervesuomi.fi, the content will be provided by a variety of parties using the ONKI ontology library, its services, and tools provided by FinnONTO. The content focus is initially on high school materials and (Finnish) history. The contents will be automatically integrated with relevant semantic content created in other

parts of FinnONTO, especially in CultureSampo. Another demonstration system called Opas will be a demonstration of how ontology techniques can be used to support help-desk services, both in indexing and in information retrieval. The system is based on an ontologized version of HKJL, the Finnish library subject heading taxonomy used by the Helsinki City Library, combined with the YSO ontology. The dataset is a collection of indexed question-answer pairs of the national public "Ask the librarian" –service (http://www.kirjastot.fi/tietopalvelu), where librarians answer to email questions of their customers. The service will be connected with CultureSampo and other semantic repositories in order to enrich the answers automatically with links to related material on the semantic web.

5. **eCommerce.** Veturi is a demonstration of semantic yellow pages (Mäkelä et al., 2005b). Here the idea was to model services and products offered by companies and the public sector with ontologies, and to provide the user with semantic searching facilities in order to ease the problem of finding matching offerings for customer needs. The dataset used was based on the yellow page data of Fonecta Ltd (over 200,000 entries) used in the commercial 020202.fi (http://020202.fi) yellow pages service, and a database of public health services.
6. **Online tools** ONKI library with its services is a demonstration of how ontologies can help in utilizing their own usage. In addition to YSO and its various vertical extension ontologies, an ontology service for the Finnish geographical names (over 800,000 entries) is being designed and implemented based on the database provided by the National Land Survey of Finland. A related effort in FinnONTO in the geography domain is to model historical changes of Finnish cities and counties (Kauppinen, Hyvönen, 2006) – a model to be used in the CultureSampo. Here the dataset used is based on a database created at the Geological Survey of Finland (GTK). We also demonstrate how distributed content creation can be facilitated by an annotation editor SAHA (http://www.seco.tkk.fi/applications/saha/). This editor, combined with the ONKI ontology library server, is in trial use in our eHealth application case Tervesuomi.fi.
7. **Other applications** According to our plan, new case studies will be started in autumn 2006 when new members join the FinnONTO project.

5. Discussion

The main idea of the FinnONTO-project is to try to build a national infrastructural foundation for the coming semantic web by establishing a large research consortium representing universities, public organizations and companies, and by working collaboratively together on the national level. The international standards and research are guiding the work, but due to language barriers and various national conventions, adaptation and application is needed. We emphasize the idea, that the technology and solutions should be transferred from the universities to the participating organizations and companies. For this purpose, many researchers from the funding organizations have been working with the research team at the university for creating the ontologies and the semantic content of the applications. Our work focuses on the applied side of the semantic web. In our view, demonstrations of good practical semantic web applications are still largely missing but needed in order to convince public organizations and companies to strive for the era of the semantic web.

Acknowledgments

Some 30 researchers working in the FinnONTO team at various periods of time have contributed to the work discussed above. FinnONTO project is currently funded by the National Technology Agency Tekes and a consortium of 29 public and private organizations.

References

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

D. Foskett (1980) Thesaurus. Encyclopaedia of Library and Information Science, Volume 30. Marcel Decker, New York, 1980.

A. Hameed, A. Preese, D. Sleeman (2004) Ontology reconciliation. In (Staab, Studer, 2004).

Markus Holi, Eero Hyvönen (2006a) Modeling Uncertainty in Semantic Web Taxonomies. In Zhongmin Ma (ed.) Soft Computing in Ontologies and Semantic Web, Springer-Verlag, 2006. Forth-coming.

Markus Holi, Eero Hyvönen (2006b) Fuzzy View-Based Semantic Search. Paper, submitted.

Eero Hyvönen, Eetu Mäkelä, Mirva Salminen, Arttu Valo, Kim Viljanen, Samppa Saarela, Miikka Junnila, Suvi Kettula (2005a) MuseumFinland - Finnish Museums on the Semantic Web. Journal of Web Semantics, vol. 3, no. 2, pp. 25, 2005.

Eero Hyvönen, Arttu Valo, Ville Komulainen, Katri Seppälä, Tomi Kauppinen, Tuukka Ruotsalo, Mirva Salminen, Anu Ylisalmi (2005b) Finnish National Ontologies for the Semantic Web - Towards a Content and Service Infrastructure. Proceedings of International Conference on Dublin Core and Metadata Applications (DC 2005), Nov, 2005.

JHS 158 (2006) Public JHS Recommendation for Geographic Metadata (in Finnish). Ministry of Internal Affairs.
http://www.jhs-suositukset.fi/intermin/hankkeet/jhs/home.nsf/pages/349B16D67CAF44FAC2256F3100228FF3

Tomi Kauppinen, Eero Hyvönen (2006) Modeling and Reasoning about Changes in Ontology Time Series. In Rajiv Kishore, Ram Ramesh, Raj Sharman (eds.) Ontologies in the Context of Information Systems. Springer-Verlag, Berlin, Dec, 2006. In press.

V. Komulainen, A. Valo, E. Hyvönen (2005) A collaborative ontology development and service framework ONKI. Paper, Helsinki Universtity of Technology, Laboratory for Media Technology, 2005.

R. L. Leskinen (ed.) (1997) Museoalan asiasanasto. Museovirasto, Helsinki, 1997.

Eetu Mäkelä, Eero Hyvönen, Samppa Saarela, Kim Viljanen (2004) OntoViews - A Tool for Creating Semantic Web Portals. Proceedings of the 3rd International Semantic Web Conference (ISWC 2004), 2004.

Eetu Mäkelä, Eero Hyvönen, Teemu Sidoroff (2005a) View-Based User Interfaces for Information Retrieval on the Semantic Web. Proceedings of the ISWC-2005 Workshop End User Semantic Web Interaction, Nov, 2005.

Eetu Mäkelä, Kim Viljanen, Petri Lindgren, Mikko Laukkanen, Eero Hyvönen (2005b) Semantic Yellow Page Service Discovery: The Veturi Portal. Poster paper, 4th International Semantic Web Conference, Nov, 2005.

Ville Komulainen, Arttu Valo, Eero Hyvönen (2005) A Collaborative Ontology Development and Service Framework ONKI. Proceeding of ESWC 2005, poster papers, 2005.

J. Sowa (2000) Knowledge representation. Logical, Philosophical, and Computational Foundations. Brooks/Cole, 2000.

S. Staab, R. Studer (eds) (2004) Handbook on ontologies. Springer-Verlag, 2004.

A. Valo, V. Komulainen, E. Hyvönen (2005) A Tool for Collaborative Ontology Development for the Semantic Web. Proceedings of International Conference on Dublin Core and Metadata Applications (DC 2005), Nov, 2005.

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