TerveSuomi.fi – A Semantic Health Portal for Citizens

Markus Holi, Petri Lindgren, Osma Suominen, Kim Viljanen, Eero Hyvönen
Semantic Computing Research Group (SeCo)
Helsinki University of Technology (TKK) and University of Helsinki
Laboratory of Media Technology, P.O. Box 5500, FI-02015 TKK
http://www.seco.tkk.fi/, email: firstname.lastname@tkk.fi

Abstract

This paper presents the on-going research for creating a prototype of a national health portal based on semantic web technologies.

1 Introduction

TerveSuomi.fi is a national health portal under construction that provides citizens with reliable and up-to-date health information created by the National Public Health Institute[1] and affiliated health organizations in Finland. The content includes metadata of thousands of web documents such as web pages, articles (e.g., "Dangers of smoking"), reports, campaign information (e.g., "National non-smoking week"), news, services, and other information related to health. Our technical challenge is to develop methods and tools 1) for creating and annotating the content by distributed information providers, 2) for aggregating the semantic content and making it semantically interoperable, and 3) for providing the citizens with easy to use semantic search and browsing of the content.

2 Ontological Metadata

The portal requires the web documents to be described in a uniform and machine-readable manner. A metadata schema specifies what information has to be expressed about each document and ontologies define the shared concepts by which the information is given. The metadata schema is expressed in RDF/XML and is based on the Dublin Core Metadata Initiative[2]. The DCMI Element Set[3] is used, along with refinements introduced in DCMI Terms[4]. We have also created some additional refinements to DCMI elements. For example, the *dc.audience* field has been refined to include *audienceAge*, *audienceSex*, *audienceFamilyStatus*, *audienceOperationalEnvironment*, *audienceProfession* and *audiencePatientGroup*.

The ontologies are based on freely available thesauri that are widely used by health organizations in Finland: 1) The Finnish version of the international Medical Subject Headings[5] (FinMeSH), which includes approximately 23 000 concepts. 2) The Finnish General Upper Ontology (YSO)[6] that includes approximately 23 000 concepts. 3) The Finnish version of the European Health Promotion Thesaurus[7] (TES). TES includes approximately 1200 concepts, which are all mapped to YSO and FinMeSH.

3 Distributed Semantic Content Creation

All content providers commit to the following rules when creating semantic content for the portal: All content must be 1) identified by unique URIs, 2) be available on the web, 3) be described with metadata, using 4) concepts from the given ontologies.

Ordinary content management systems (CMS) do not currently support ontological metadata, only freetext and controlled vocabularies. In our case, four different ways for creating and storing ontological metadata are supported:

1) Semantic tags can be used instead of keywords in the existing CMSs for storing ontological concepts. A semantic tag is a human readable keyword (may contain spaces etc.) combined with a namespace to help to identify the intended concept in a certain ontology, e.g., "mesh:Tobacco Smoke Pollution". When using semantic tags, the CMS does not have to support ontologies or URIs. The mapping of the semantic tags to ontological concepts is done outside of the CMS. The CMS must be able to export the metadata, e.g., as RDF.

2) Ontological functionalities such as concept search and storing of ontological metadata can be added to an existing CMS. The ontology server ONKI provides, e.g., a web service interface, an HTML/Ajax component which can be included in the editing page of a CMS, and an HTML-based user interface for looking up concepts [Komulainen *et al.*, 2005]. The ontology files can also be downloaded from ONKI, if the content provider chooses to implement the ontological services inside the CMS. In all cases, the CMS must store the concept URIs as a part of the metadata. The CMS must be able to export the metadata, preferably as RDF.

3) The semantic annotation editor SAHA can be used for storing document metadata either combined with an existing

[1] http://www.ktl.fi
[2] http://dublincore.org/
[3] http://dublincore.org/documents/dces/
[4] http://dublincore.org/documents/dcmi-terms/
[5] http://www.nlm.nih.gov/mesh/
[6] http://www.seco.tkk.fi/ontologies/yso/
[7] http://www.hpmulti.net/

CMS or as an independent system [Valkeapää and Hyvönen, 2006]. When combined with a CMS, SAHA stores only those metadata fields that do not exist in the CMS. The metadata from SAHA and the CMS is combined before using it in the health portal. SAHA can also be used for storing all metadata about the documents. This solution can be used if the content provider does not have a CMS, the CMS cannot be used for storing metadata or the content is created by a third party. SAHA supports exporting the metadata in RDF.

4) Semantically rich content metadata may be transformed automatically to the ontological metadata format used in the health portal. The transformation is done in two steps: First the ontologies, classifications, thesauri etc. are mapped to the health portal ontologies. Then the metadata is transformed using these mappings. An initial version of a semiautomatic mapping tool which maps concepts based on term similarity and ontological structure has been created.

4 Fuzzy User-centric Search Views

The user interface of the portal is based on the view-based semantic search paradigm [Hyvönen *et al.*, 2004]. A special problem in TerveSuomi.fi is that the ontologies used for annotating the health content are intended for health information professionals to use, whereas the portal is mainly targeted at the general public. This means that the annotation ontologies and their concept hierarchies cannot be used directly for querying in navigational facets, as in semantic portals such as MuseumFinland[8]. Instead, we are constructing new, citizen-centric ontologies and views (facets) by using the technique of card sorting [Rugg and McGeorge, 1997]. After this, fuzzy mappings [Holi and Hyvönen, 2006] are used to map the query categories to the annotation concepts, and to rank search results according to their relevance to the query. The primary navigational facets are Topic, Life event, Group of people and Body part. The interface also has secondary (drop-down) facets for specifying Document type, Publisher, Publication year and Audience.

An illustration of the planned user interface is shown in Figure 1. In addition to view-based search, the portal provides recommended links based on ontological knowledge (e.g. "smoking is a risk factor for lung cancer"), graphical concept maps, and an alphabetical index of concepts.

5 Conclusions and Future Work

The main contributions of this work are the use of semantic web technologies for 1) distributed content creation and aggregation, 2) building user-centric view-based navigation and 3) fuzzy mapping of query view categories to annotation ontologies.

We are currently also investigating how ontologies could be used to model health care services. In the future, the portal may be extended to incorporate access to personal medical records and health care services.

This work is part of the FinnONTO project[9], funded mainly by the National Funding Agency for Technology Innovation (Tekes) and the Ministry of Social Affairs and Health. The TerveSuomi.fi project is co-ordinated by the National Public Health Institute in Finland (Project Coordinator Eija Hukka).

[8] http://www.museosuomi.fi

[9] http://www.seco.tkk.fi/projects/finnonto/

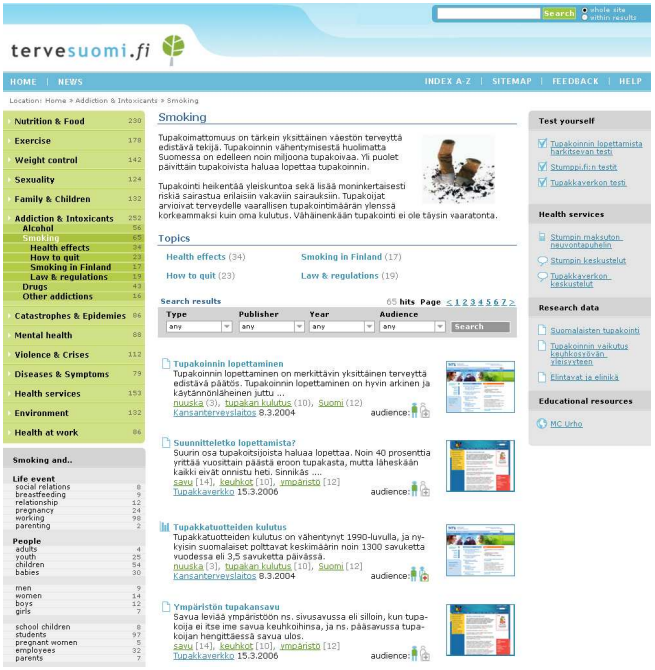

Figure 1: TerveSuomi semantic user interface

References

[Holi and Hyvönen, 2006] Markus Holi and Eero Hyvönen. Fuzzy view-based semantic search. In *Proc. of the 1st Asian Semantic Web Conference (ASWC2006), Beijing, China*. Springer-Verlag, September 3-7 2006. To be published.

[Hyvönen *et al.*, 2004] Eero Hyvönen, Samppa Saarela, and Kim Viljanen. Application of ontology techniques to view-based semantic search and browsing. In *The Semantic Web: Research and Applications. Proc. of the 1st European Semantic Web Symposium (ESWS 2004)*, 2004.

[Komulainen *et al.*, 2005] Ville Komulainen, Arttu Valo, and Eero Hyvönen. A tool for collaborative ontology development for the semantic web. In *Proc. of the International Conference on Dublin Core and Metadata Applications (DC 2005)*, Nov 2005.

[Rugg and McGeorge, 1997] Gordon Rugg and Peter McGeorge. The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. *Expert Systems*, 14(2):80–93, 1997.

[Valkeapää and Hyvönen, 2006] Onni Valkeapää and Eero Hyvönen. A browser-based semantic annotation tool for distributed content creation, June 16 2006. Poster paper, submitted for evaluation to 1st Asian Semantic Web Conference (ASWC).