

# Semantic Faceted Search in a Citizens' Health Portal

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](mailto:firstname.lastname@tkk.fi)

## Abstract

This paper presents TerveSuomi.fi, a prototype of a national semantic health portal in Finland. TerveSuomi.fi aggregates ontological metadata created by various Finnish health organizations in a distributed environment. The portal provides a faceted search user interface created from the perspective of ordinary citizens and cross-links documents from multiple sources with recommendations based on ontological knowledge.

## 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 Health Institute<sup>1</sup> 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 based on the the Dublin Core Element Set<sup>2</sup>, along with refinements introduced in DCMI Terms<sup>3</sup>. We have also created some additional refinements to DCMI elements. For example, the *dc:type* field has been refined with a *genre* field, with values such as *News item*, *Organizational information* and *Research*.

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<sup>4</sup> (FinMeSH), which includes approximately 23 000 concepts. 2) The Finnish General Upper Ontology (YSO)<sup>5</sup> that includes approximately 23 000 concepts. 3) The Finnish version of the European Health Promotion Thesaurus<sup>6</sup> (HP-MULTI). HPMULTI 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 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) The ontologies can be used like traditional controlled vocabularies in existing CMSs as long as the source vocabulary for each concept is indicated. Human-readable concept labels are used in indexing, and the CMS does not have to support ontologies or URIs. The mapping of the labels to concept URIs is done outside of the CMS. The CMS must be able to export the metadata as RDF or HTML meta tags.

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 CMS or as an independent system [Valkeapää *et al.*, 2007]. This solution can be used if the content provider does not have a CMS, the CMS cannot be used for storing metadata

<sup>1</sup><http://www.ktl.fi>

<sup>2</sup><http://dublincore.org/documents/dces/>

<sup>3</sup><http://dublincore.org/documents/dcml-terms/>

<sup>4</sup><http://www.nlm.nih.gov/mesh/>

<sup>5</sup><http://www.seco.tkk.fi/ontologies/ys/>

<sup>6</sup><http://www.hpmulti.net/>

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.

## 4 User-centric Faceted Search

The user interface of the portal (Figure 1) 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<sup>7</sup>. Instead, we have constructed new, citizen-centric facets and mapped these to the underlying ontologies [Suominen *et al.*, 2007]. 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.

In addition to faceted search (Figure 2), the portal provides recommended links based on ontological knowledge (e.g. "smoking is a risk factor for lung cancer"), a concept browser (Figure 3), and an alphabetical index of concepts (Figure 4).



Figure 1: TerveSuomi portal

<sup>7</sup><http://www.museosuomi.fi>

## 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 faceted navigation and 3) mapping of facet 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.

## 6 Acknowledgments

This work is part of the FinnONTO project<sup>8</sup>, 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 Health Institute in Finland (Project Coordinator Eija Hukka). We thank Markus Holli and Petri Lindgren for their input to the work reported in this paper.

## References

- [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.
- [Suominen *et al.*, 2007] Osma Suominen, Kim Viljanen, and Eero Hyvönen. User-centric faceted search for semantic portals. In *Proc. of ESWC 2007, Innsbruck, Austria*, June 3–7 2007.
- [Valkeapää *et al.*, 2007] Onni Valkeapää, Olli Alm, and Eero Hyvönen. Efficient content creation on the semantic web using metadata schemas with domain ontology services. In *Proc. of ESWC 2007, Innsbruck, Austria*, June 3–7 2007. System demonstration paper.

<sup>8</sup><http://www.seco.tkk.fi/projects/finnonto/>

Secondary facets: Type, Publisher, Publication year, Audience

Topic facet  
selected category: Diet  
(Ruokavalio)

Recommended  
links

Life event facet  
selected category:  
Pregnancy (raskaus)

Group of people  
facet

Search results  
about diet and pregnancy

The screenshot shows the website interface for 'terveysuomi.fi'. The main content area displays search results for 'Ruokavalio ja raskaus'. The results are filtered by 'Diet (Ruokavalio)' and 'Pregnancy (raskaus)'. The search results list several articles, including 'Raskaana olevien ja imettävien äitien rokottaminen', 'Taulukko 7. Tilanteita, jotka eivät ole vasta-aiheisia rokottamiselle', 'Ravitsemus', 'Raskauden aikainen ruokavalio / Ravitsemus', and 'Sosiaali- ja terveysministeriön antama D-vitamiinivalmisteiden käyttösuositus'. The interface includes a search bar at the top, a navigation menu, and a sidebar with various facets and filters. The facets include 'Topic facet', 'Life event facet', and 'Group of people facet'. The search results are displayed in a list format with titles, brief descriptions, and links to the full articles.

Figure 2: Semantic faceted search

Sijainti: Käsitteet > Tobacco

- ▶ Ravinto ja ruoka
- ▶ Liikunta
- ▶ Painonhallinta
- ▶ Seksuaalisuus
- ▶ Perhe ja lapset
- ▶ Päihteet ja nautintoaineet 1
- ▶ Katastrofit
- ▶ Epidemiat ja tarttuvat taudit
- ▶ Mielenterveys, uni ja jaksaminen
- ▶ Väkivalta ja kriisit
- ▶ Sairaudet ja oireet
- ▶ Tapaturmat ja ensiapu
- ▶ Mistä apua? 1
- ▶ Ympäristö
- ▶ Työkyky ja työterveys

**Tobacco ja...**

**Elämäntilanne**  
eläkkeelle siirtyminen  
ikäntyminen  
imetys  
isvvs

**Käsite: Tobacco**

**Päätermi** Tobacco

**Muut termit** Nicotiana

**Laajemmat käsitteet** [\(nautintoaineet\)](#) (2)  
[Solanaceae](#) (1)

**Suppeammat käsitteet** [Tobacco](#), [Smokeless](#) (0)

**Rinnakkaiset käsitteet** [\(tupakkatuotteet\)](#) (0)  
[Nicotine](#) (0)  
[Nicotine dependency](#) (0)  
[Smoking](#) (0)  
[Tobacco consumption](#) (0)

**Vastaavat käsitteet** <http://www.yso.fi/onto/hpmulti/T01172>  
<http://www.yso.fi/onto/mesh/D014026>  
[http://yso.fi/YSO#tupakka\\_2](http://yso.fi/YSO#tupakka_2)

**Hakutulokset**

1 kpl Sivut 1

Tyyppi	Julkaisija	Vuosi	Yleisö
<input type="text" value="mikä tahansa"/>	<input type="text" value="mikä tahansa"/>	<input type="text" value="mikä tahansa"/>	<input type="text" value="mikä tahansa"/>

Rajaa hakua

 [Decree on Measures to Reduce Tobacco Smoking \(päihteet\), \(tupakka\), \(yrityspalvelut\), Health, Tobacco FINLEX / Ministry of Social Affairs and Health 1977 → \(asiantuntijat\)](#)






1 kpl Sivut 1




**Testaa itsesi**

- [Tupakoinnin lopettamista harkitsevan testi](#)
- [Stunppi.fi:n testi](#)
- [Tupakkaverkon testi](#)

**Terveyspalveluita**

-  [Stunpin maksuton neuvontapuhelin](#)
-  [Stunpin keskustelut](#)
-  [Tupakkaverkon keskustelut](#)

**Tutkimustietoa**

-  [Suomalaisten tupakointi](#)
-  [Tupakoinnin vaikutus keuhkosyövän yleisyyteen](#)
-  [Elintavat ja elinikä](#)

**Opetusmateriaalia**



-  [MC Urho](#)

Figure 3: Concept browsing

tervesuomi.fi 

ETUSIVU | UUTISET | HAKEMISTO A-Ö | SIMUKARTTA | PALAUTE | OHJE

Sijainti: Hakemisto > U

**Hakemisto: U**

[0-9](#) [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) **U** [V](#) [W](#) [X](#) [Y](#) [Z](#) [Å](#) [Ä](#) [Ö](#)

U	ulkomaat (45)	uranvalinta (2)
kts. uraani (5)	ulkosuomalaiset (8)	<b>uranvalinta</b>
UKK-kävelytesti (14)	<b>ulkukset</b>	kts. ammatinvalinta (4)
UV	kts. haavat (3)	urheilijat (5)
kts. ultraviolettisäteily (1)	ulosotto (16)	urheilu (87)
UV-säde	ulosteet (2)	<b>urheilu-doping</b>
kts. ultraviolettisäteily (1)	<b>ulseraatio</b>	kts. (Doping in Sports) (1)
UV-säteily	kts. haavat (3)	urheilukilpailut (1)
kts. ultraviolettisäteily (1)	<b>ultraviolettisäde</b>	urheilulajit (2)
uhanalaiset lajit (2)	kts. ultraviolettisäteily (1)	urheiluvammat (2)
uimahallit (1)	ultraviolettisäteily (1)	<b>urtikaria</b>
uimarannat (1)	ummetus (2)	kts. nokkosihottuma (5)
uimataito (1)	<b>unenaikaiset hengityshäiriöt</b>	<b>urtikariapaukama</b>
uimavesi (3)	kts. uniapnea-oireyhtymä (1)	kts. nokkosihottuma (5)
uiminen	unettomuus (3)	uskomukset (1)
kts. uinti (2)	uni (3)	uskonnot (10)
uinti (2)	uni (1)	uskonto (11)
ulkoilu (16)	uniapnea-oireyhtymä (1)	uupumus (4)
ulkoilualueet (3)	<b>uniapnea-syndroomat</b>	<b>uusiutuva rasitusvamma</b>
<b>ulkoinen laskentatoimi</b>	kts. uniapnea-oireyhtymä (1)	kts. rasitusvammat (3)
kts. kirjanpito (16)	unihäiriöt (6)	uusperheet (2)
ulkomaalaiset (30)	unilääkkeet (1)	uutiset (22)
ulkomaankauppa (37)	unkarilaiset (1)	
ulkomaanmatkailu (4)	uraani (5)	
<b>ulkomaantyö</b>		
kts. työskentely ulkomailla (23)		

**Katso myös**

- [Sivukartta](#)

Figure 4: Alphabetical concept index