



HELSINKI UNIVERSITY OF TECHNOLOGY  
Media Technology

# HealthFinland (TerveSuomi)

## Finnish Health Information on the Semantic Web

30 November 2007

**Osma Suominen**, Kim Viljanen, Eero Hyvönen

Semantic Computing Research Group (SeCo)

Helsinki University of Technology (TKK),

Laboratory of Media Technology

and

University of Helsinki, Department of Computer Science

<http://www.seco.tkk.fi/>



UNIVERSITY OF HELSINKI



SeCo  
semantic computing

# Outline of Talk



HELSINKI UNIVERSITY OF TECHNOLOGY  
Media Technology

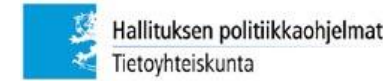
- | Goals of HealthFinland
- | Challenges
- | Content creation
- | KOS in HealthFinland
- | User interface and navigation
  - demonstration of portal
- | Evaluation of results

# HealthFinland portal



HELSINKI UNIVERSITY OF TECHNOLOGY  
Media Technology

- | national health promotion portal for citizens [1]
- | SW technologies: ontologies, metadata...
  - content aggregated from many websites into a single portal
- | diet, exercise, (non)smoking, healthy living...



[1] Hyvönen et al: **HealthFinland - Finnish Health Information on the Semantic Web**. Proc. ISWC 2007, Busan , Korea, Nov 2007



<http://www.seco.tkk.fi>



# Goals of HealthFinland: Citizens



HELSINKI UNIVERSITY OF TECHNOLOGY  
Media Technology

- | **Global view** to health information from different organizations
- | **Aggregated view** to information from different organizations
- | **Semantic search and browsing**
  - faceted browsing using user-centric categorizations
  - string search based on ontologies
  - semantic recommendations of related information

## Goals of HealthFinland: Publishers



HELSINKI UNIVERSITY OF TECHNOLOGY  
Media Technology

- | Rationalizing content creation by **eliminating redundancy**
- | **Enriching content** with other providers' content
- | Automatic and **dynamic content linking**
- | Reusing global services cost-efficiently as Web 2.0 mash-ups
- | Using centralized ontology **services for indexing**

# Challenges

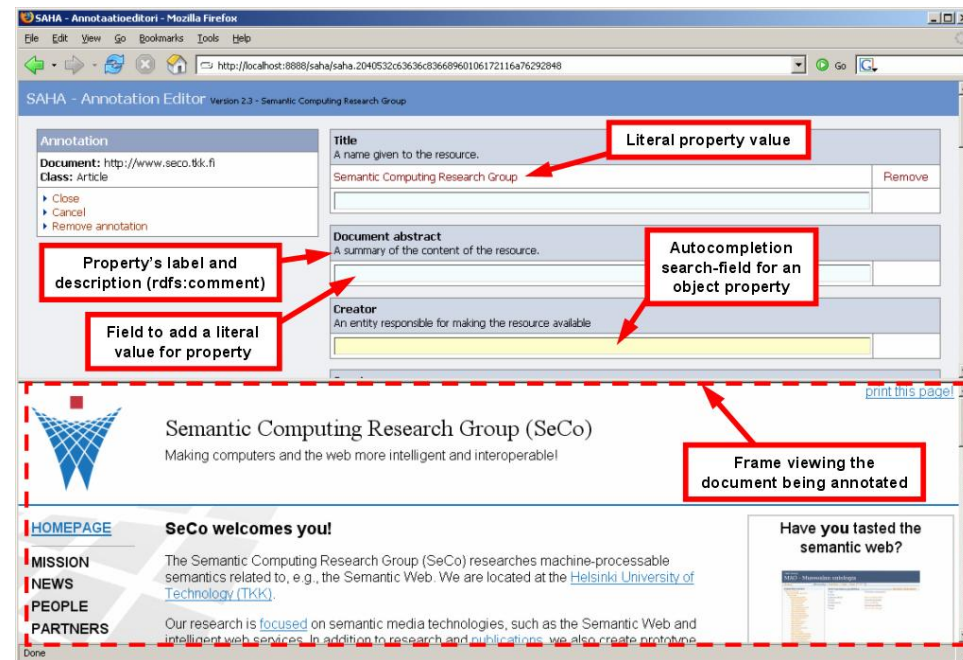


- | create a compelling user experience for the general public
  - usability is not just a surface feature!
  - solve actual problems users have – in an intuitive way
- | tri-lingual portal and KOS
  - Finnish, Swedish, English
- | interoperable
  - with existing KOS
  - legacy metadata, other document repositories and portals
  - current and future content management systems
- | need to gain a critical mass of information, publishers, users
- | need processes for content creation, quality control, dealing with problems, KOS updates and maintenance

# Content Creation Tools



- | Content harvester collects metadata from HTML pages and RDF sources
- | Metadata validator with feedback reports
- | SAHA Annotation editor
- | ONKI Ontology Server



# Metadata Schema in HealthFinland



**Table 1.** HEALTHFINLAND Metadata Schema. Obligatory fields are marked in **bold**. Cardinalities are presented in the column C.

	Name	QName	C	Value type	Value range
General metadata	<b>Identifier</b>	<b>dc:identifier</b>	1	URI	
	Locator	ts:url	0..1	URL	
	<b>Title</b>	<b>dc:title</b>	1 <sup>a</sup>	Free text	Non-empty string.
	<b>Abstract</b>	<b>dcterms:abstract</b>	1 <sup>a</sup>	Free text	Non-empty string.
	<b>Language</b>	<b>dc:language</b>	1..*	String	RFC 3066
	<b>Publication time</b>	<b>dcterms:issued</b>	1	String	W3CDTF (ISO 8601)
	Acceptance time	dcterms:dateAccepted	0..*	String	W3CDTF (ISO 8601)
	Modification time	dcterms:modified	0..*	String	W3CDTF (ISO 8601)
	<b>Publisher</b>	<b>dc:publisher</b>	1..*	Instance	foaf:Organization
	Creator	dc:creator	0..*	Instance	foaf:Organization, foaf:Person or foaf:Group
Content classification	<b>Subject</b>	<b>dc:subject</b>	1..*	Concept	YSO, MeSH and HPMulti Ontologies
	<b>Audience</b>	<b>dcterms:audience</b>	1..*	Concept	Audience Ontology
	<b>Genre</b>	<b>ts:genre</b>	1..*	Concept	Genre Ontology
	<b>Presentation type</b>	<b>dc:type</b>	1..*	Concept	DCMI Type vocabulary
	<b>Format</b>	<b>dc:format</b>	1	String	IANA MIME types
	<b>Medium</b>	<b>dcterms:medium</b>	1	Concept	Medium Ontology
	Spatial coverage	dcterms:spatial	0..*	String or concept	DCMI Point, DCMI Box or Location Ontology
	Temporal coverage	dcterms:temporal	0..*	String or concept	W3CDTF, DCMI Period or Time Ontology
Relations	Part of	dcterms:isPartOf	0..*	Document	URI
	Rights	dc:rights	0..*	Free text or document	URI or textual description
	Source	dc:source	0..*	Free text or document	URI (e.g., ISBN) or bibliographical reference
	Reference	dcterms:references	0..*	Free text or document	URI (e.g., ISBN) or bibliographical reference
	Translation of	ts:isTranslationOf	0..*	Document	URI
Format of	dcterms:isFormatOf	0..*	Document	URI	

<sup>a</sup> Multilingual values are allowed, but only one value in each language.



# KOS in HealthFinland



HELSINKI UNIVERSITY OF TECHNOLOGY  
Media Technology

## Source KOS

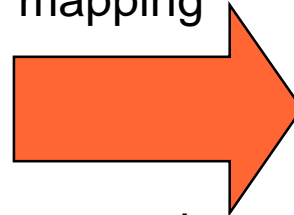
**YSO**  
>20000 concepts  
fi, en, sv  
OWL ontology  
created by FinnONTO (from YSA)  
General knowledge

**MeSH**  
>20000 concepts  
en, fi, sv ...  
SKOS thesaurus  
created by US NLM  
Medical domain (for experts)

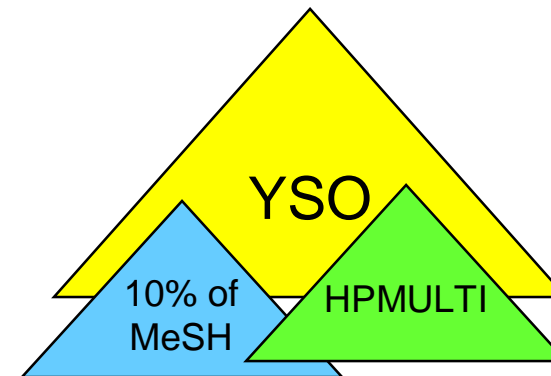
**HPMULTI**  
1200 concepts  
en, fi, sv ...  
SKOS thesaurus  
created by EU project  
Health promotion specific

+ some organization-specific thesauri  
(TTL, STAMETA, Suomi.fi ...)

automatic  
term-based  
mapping



manual  
ontology  
engineering



## HealthFinland Health Promotion Ontology

(all stored and maintained  
on Onki ontology server)



UNIVERSITY OF HELSINKI

<http://www.seco.tkk.fi>





- | **Problem:**
  - complex ontologies not suitable for human consumption**
    - made for a different purpose (e.g. indexing scientific articles)
    - expert terminology
    - unintuitive hierarchies and groupings
    - everything in a big hierarchy – not facet-based
  
- | **need to build navigation structures that users of the site will understand [2]**
  
- | **solved using an approach [3] based on card sorting**

[2] Pollit: **The key role of classification and indexing in view-based searching.**

Technical report. University of Huddersfield, UK (1998)

[3] Suominen et al: **User-centric faceted search for semantic portals.**

Proc. ESWC 2007, Innsbruck, Austria, Jun 2007

# Example of problem



- | My 70 year old aunt is depressed. How can I find information about mental illnesses in elderly people?

## MeSH Tree Structures - 2007

1. **+** Anatomy [A]
2. **+** Organisms [B]
3. **+** Diseases [C]
4. **+** Chemicals and Drugs [D]
5. **+** Analytical, Diagnostic and Therapeutic Techniques and Equipment [E]
6. **+** Psychiatry and Psychology [F]
7. **+** Biological Sciences [G]
8. **+** Natural Sciences [H]
9. **+** Anthropology, Education, Sociology and Social Phenomena [I]
10. **+** Technology, Industry, Agriculture [J]
11. **+** Humanities [K]
12. **+** Information Science [L]
13. **+** Named Groups [M]
14. **+** Health Care [N]
15. **+** Publication Characteristics [V]
16. **+** Geographicals [Z]



# Demonstration of HealthFinland



HELSINKI UNIVERSITY OF TECHNOLOGY  
Media Technology

**terveysuomi.fi**

Kansalaisen portti luotettavaan terveystietoon

suomeksi | på svenska | in English

ETUSIVU | UUTISET

HAKEMISTO A-O | SIVUKARITTA | PALAUTE | OHJE

Sijainti: Etusivu

Terveysuomi.fi on väylä luotettavaan ja monipuoliseen terveystietoon. Sisällöstä vastaavat terveysalan tutkimus- ja asiantuntijalaitokset, viranomaiset ja alan järjestöt.

**Aihealueet**

- Food and nutrition**  
Diet and eating habits, Foodstuffs, Nutrients...
- Exercise**  
Everyday exercise, Fitness and physical activity, Leisure time exercise...
- Weight control**  
Body mass index, Body weight, Slimming...
- Sexuality**  
Birth control, Sexually transmitted diseases...
- Family and children**  
Delivery, Family problems, Pregnancy...
- Intoxicants**  
Alcohol, Drugs, Smoking...
- Catastrophes**  
Chemical accidents...
- Epidemics and infectious diseases**  
Bacteria, Epidemic prevention, Vaccines...
- Mental health and sleep**  
Depression, Self-destruction, Sleep and sleep disorders...
- Violence and crises**  
Marginalisation, Violence...
- Conditions and symptoms**  
Conditions, Symptoms, Treatment methods...
- Accidents and first aid**  
Falling, Stays, Suicide and injuries...
- Findings**  
Elderly care, Health services, Pharmacies...
- Environment**  
Air, Harmful substances, Residential environment...
- Occupational health**  
koulutus, toimintakyky, työ...

**Suurimmat aiheet**

Air Pollutants Nutrition Older people Patients People with psychiatric disorders Women Young adults adults bipolar disorder children **depression** epidemiology fatness mental disorders mood nurse-patient relationship publications **research** risk factors **schizophrenia** siblings suicide **suicide attempts** young people

**Hakemisto**

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

**Uutiset**

- Kansanterveyslaitosten kansainvälinen järjestö perustettu**  
3.2.2006  
Maailman kansanterveyslaitosten pääjohtajat kokoontuivat 29.1.-1.2.2006 Rio de Janeirossa.
- Veroale ei lisännyt eläkeikäisten alkoholinkäyttöä**  
2.2.2006  
KTL:n tutkimus ei havainnut merkittävää alkoholinkäytön lisääntymistä vuosina 2003-2005.
- Wallenberg-säätiö panostaa pohjoismaiseen tautigenetiikkaan**  
2.2.2006  
Nordic Centre of Excellence in Disease Genetics pohjoismaisen geenitutkimuksen huippuyksikkö.
- Väitös 10.2: Kuntoilu, lihavuus ja sydäntautien vaaratekijät**  
30.1.2006  
Väittelijänä liikuntatieteiden maisteri Katja Borodulin.
- 200 000 suomalaisella piilodiabetes**  
19.1.2006  
Tyyppin 2 diabetes ja sen esiasteet oletettavain yleisempiä.

Lisää uutisia...

Uutiset RSS-virtana  
Mikä on RSS?

© Julkaisija: terveysuomi.fi 2006 Tietoja sivustosta



UNIVERSITY OF HELSINKI

http://www.seco.tkk.fi



SeCo  
semantic computing

# Solution approach



HELSINKI UNIVERSITY OF TECHNOLOGY  
Media Technology

- | faceted browsing very useful and usable in earlier portals
  - Flamenco [4], SWED [5], MuseumFinland [6] ...
- | so, need to build intuitive facets and categories for the site
- | to find out what works for users, do real user research [7]
- | when you know what works for users, handle the technicalities
  - mapping user facets to ontologies

[4] Hearst et al: **Finding the flow in web site search.** CACM 45(9) 2002

[5] <http://www.swed.org.uk>

[6] Hyvönen et al: **MuseumFinland – Finnish museums on the Semantic Web.**  
Journal of Web Semantics 3(2) 2005

[7] Rugg & McGeorge: **The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts.** Expert systems 14(2) 1997

# Card sorting



- | method to find out how users conceptualize the information space (i.e. how they group things in their head)
- | often used to build website navigation (information architecture)
  - what should the main sections of the site be
  - how should they be named
  - what things go where
  - what things belong together
- | idea: print a stack of cards with names of documents, let users sort them into piles and give names to the piles
  - make notes
- | repeat with several users, try to find common patterns
- | raw output: sets of labeled piles of cards
- | easy to do, very effective, enlightening for a designer



VÄESTÖRYHMÄT

OMAT ONGELMAT /  
LIEVEILMIÖT

- per...
- vanhemmat 158
- koululaiset 58
- kehitysvammaiset 52
- maahanmuuttajat 82
- työikäiset 148
- keski-ikäiset 54
- eläkeläiset
- tytöt
- nais
- vanhuk
- lapset
- nuoret
- vastasyntyne

- huumeet 34
- tupakointi 145
- päihteet 108
- alkoholi 5
- väkivalta 172
- itsemurha 44
- tupakka 144

# Facets and categories



## Topic

- Addiction & Intoxicants  
e.g., smoking, alcohol
- Catastrophes
- Diseases & Symptoms
- Environment
- Epidemics
- Exercise
- Family & Children
- Health at Work
- Health Services
- Mental Health
- Nutrition & Food
  - Food Products
  - Diet
  - Nutrients
- Sexuality
- Weight Control
- Violence & Crises

## Group of People

- Age group  
e.g., adult, child, baby, elderly
- Gender  
men, women, girls, boys
- Profession / Role
- Special Groups

## Life Event

Concepts, e.g., growing up, death, pregnancy

## Body Part

Concepts, e.g., teeth, nervous system, heart



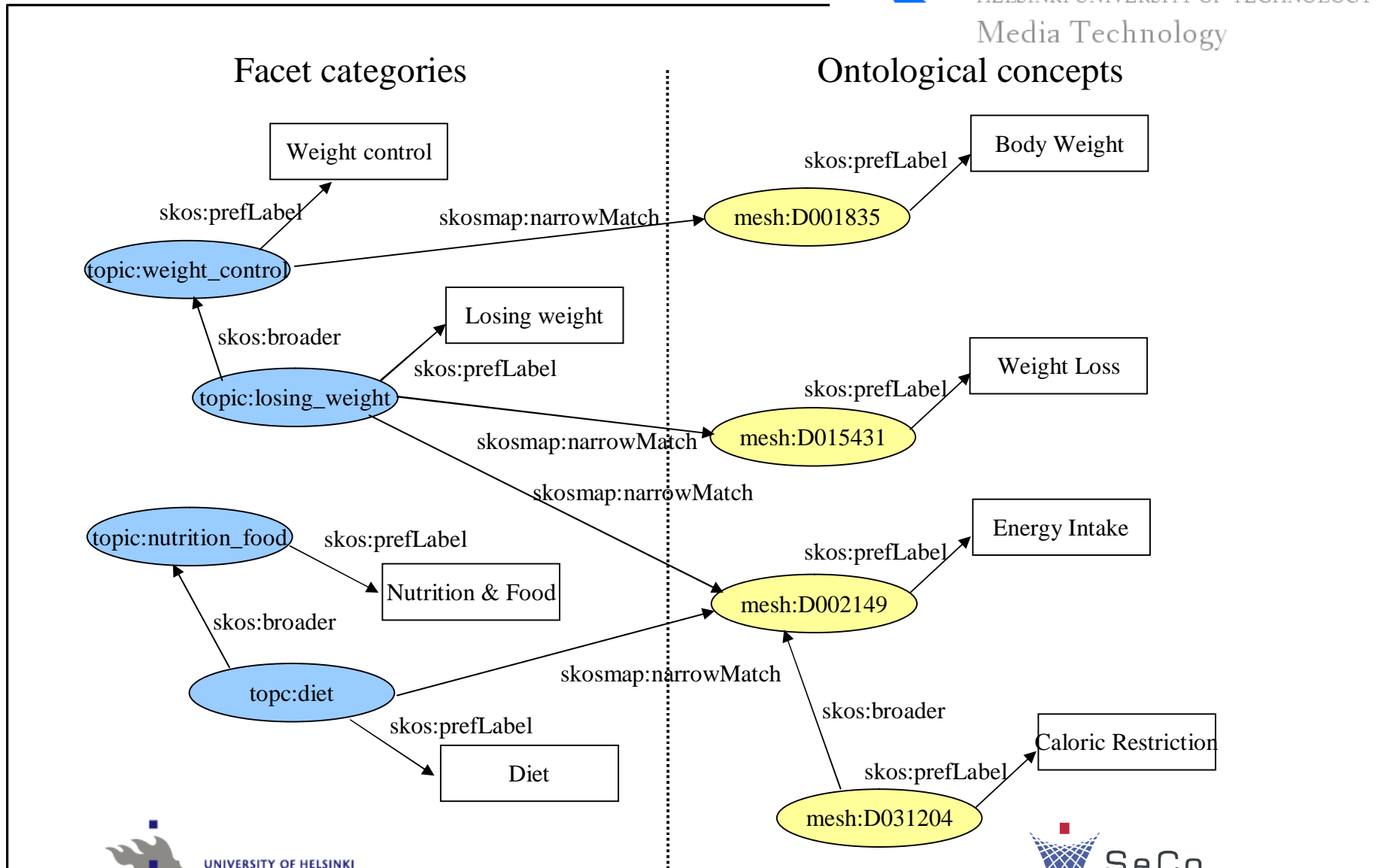


# Mapping facets to ontology



- | each category contains a set of concepts
- | category taxonomy represented using SKOS vocabulary
  - labels: skos:prefLabel
  - hierarchy: skos:broader
- | mappings to ontology using SKOS Mapping
  - skosmap: exactMatch, skosmap:narrowMatch
- | result: categories contain documents
  - annotated with concepts contained in the category
    - » or subconcepts (rdfs:subClassOf or skos:broader)
  - from all subcategories, recursively
    - » creating a proper subsumption hierarchy

# Example mappings



# Evaluation



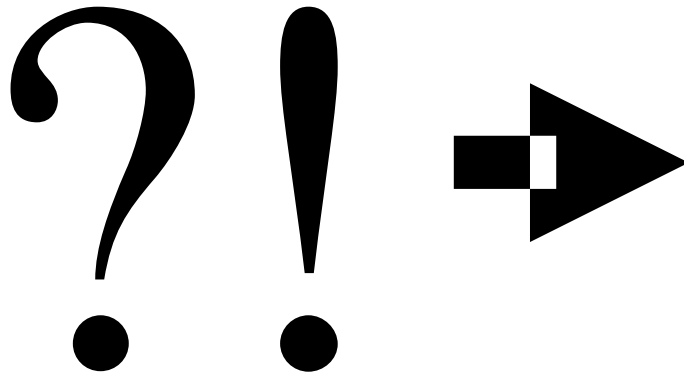
- | closed card sorting session to test intuitiveness of facets
  - given the final facets, do test subjects place a set of concepts in the intended categories?
  - promising results (but only 2 subjects)
- | review of facets by domain experts: some problems corrected
  - too much lumping gives wrong message to users
  - omission of important topics
- | prototype portal to test whether the approach can actually be implemented
- | currently performing 3<sup>rd</sup> phase user testing with prototype
  - so far the problems found have not concerned the categorizations
  - more user tests under way (in the following 2 weeks)
  - results will be published in academic conferences, journals etc.
- | prototype will go live in early 2008, production use in 2008-2009



# Thanks!



HELSINKI UNIVERSITY OF TECHNOLOGY  
Media Technology



Semantic Computing Research Group (SeCo)  
Making computers and the web more intelligent and interoperable!

**SeCo welcomes you!**  
The Semantic Computing Research Group (SeCo) researches machine-processable semantics related to, e.g., the Semantic Web. We are located at the [Helsinki University of Technology \(TKK\)](#).

**»»» Työtä tarjolla! SeCo is hiring! «««**

Our research is **focused** on semantic media technologies, such as the Semantic Web and intelligent web services. In addition to research and **publications**, we also create prototype applications that demonstrate the new possibilities of semantic technologies, such as **semantic portals for end-users**, **semantic infrastructural services**, and **ontologies and tools** for creating semantic applications.

**Selected SeCo Applications for End Users**

- [MuseumFinland Try It!](#)
- [Orava Try It!](#)
- [Kulttuurisampo](#)

**Latest News**

FinnONTO Awarded at FENIX Final Seminar, FinnONTO palkittiin FENIX-ohjelman loppuseminaarissa  
FinnONTO won the best project award in category of knowledge...  
25.4.2007 12:58 by jhkuri

Several results were published at FENIX - Interactive Computing 2003-2007 Final Seminar, SeCo:n tuloksia julkistettu FENIX-tekniologiaohjelman loppuseminaarissa  
SeCo released a first version of YSO at FENIX Results Promotion J...  
23.4.2007 19:08 by jhkuri

[www.seco.tkk.fi](http://www.seco.tkk.fi)

[osma.suominen@tkk.fi](mailto:osma.suominen@tkk.fi)

Hyvönen et al: **HealthFinland - Finnish Health Information on the Semantic Web.**

Proc. ISWC 2007, Busan , Korea, Nov 2007

Suominen et al: **User-centric faceted search for semantic portals.**

Proc. ESWC 2007, Innsbruck, Austria, Jun 2007



<http://www.seco.tkk.fi>

