



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

TerveSuomi

– terveystieto semanttisessa webissä

Eero Hyvönen

Professori, tutkimusjohtaja
Semanttisen laskennan tutkimusryhmä (SeCo)
Teknillinen korkeakoulu, mediatekniikan laitos (TKK)
Helsingin yliopisto, tietojenkäsittelytieteen laitos
<http://www.seco.tkk.fi/>
eero.hyvonen@tkk.fi



UNIVERSITY OF HELSINKI



SeCo



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

Tervetuloa Teknillisen korkeakoulun Mediatekniikan laitokselle



UNIVERSITY OF HELSINKI



SeCo

Miksi tämä tilaisuus järjestettiin?



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Kansallisesti merkittävän uuden portaalin TerveSuomi prototyyppi on valmis
 - **Kansalaisille** uusi tärkeä kansallinen palvelu
 - **Sisällön tuottajille** uusi julkaisukanava
 - **Muille verkkosovelluksille** hyödyllinen sisältölähde
- Uusi teknologinen ratkaisumalli: semanttinen web
 - Paremmat ”älykkäät” palvelut kansalaisille
 - Sisältötyön parempi koordinointi ja säästöt
 - Kustannustehokas sisältöjen koonti ja uusiokäyttö
 - Järjestelmä on ensimmäinen laatuaan maailmassa
- Demonstraatio Kansallisesta ontologiakirjastopalvelusta ONKI



Hankkeen toteutus: taustaprojektit



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Kansalaisen terveystietoportti 2005-
 - Kansanterveyslaitos johtaa konsortiota
 - Sosiaali- ja terveysministeriö rahoittaa
 - Ongelman omistaja ja sisällön tuottaja
- FinnONTO 2003-2008, 2009-2010
Suomalaiset semanttisen webin ontologiat
 - Teknillinen korkeakoulu ja Helsingin yliopisto vetävät
 - Semanttisen laskennan tutkimusryhmä SeCo
 - Tekes ja 38 muuta organisaatiota rahoittavat
 - Ratkaisukonseptin ja prototyypin kehittäminen





HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Terveyden edistämisen keskus
 - Järjestöjen terveystaali -hanke
 - RAY rahoittaa
 - Hanke integroituu TerveSuomen kanssa
- Aikataulu
 - Projektin käynnistyminen 2005
 - TKK/HY julkaisee prototyypin syyskuussa 2008
 - KTL julkaisee lopullisen tuotteen helmikuussa 2009



UNIVERSITY OF HELSINKI



SeCo

TerveSuomi: yhteistyökumppaneita



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology



SOSIAALI- JA TERVEYSMINISTERIÖ



Sosiaali- ja terveysalan tutkimus- ja kehittämiskeskus



TERVEYDEN EDISTÄMISEN KESKUS
CENTRE FOR HEALTH PROMOTION



Hallituksen politiikkaohjelmat
Tietoyhteiskunta



Työterveyslaitos

suomi.fi

SITRA

Kela



UNIVERSITY OF HELSINKI



SeCo

FinnONTO: yhteistyökumppaneita



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology



Sisällön tuottajat ja aineistot TerveSuomi-prototyypissä



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Pilotin sisällöntuottajat
 - 8 eri tuottajaa
- Pilotin sisällöt
 - 15 aineistotyyppiä
 - 8 aineistolähdettä
 - 1700 sisältökohdetta

Aineistotyypin mukaan	
aiankohtaista	32
lehtiartikkeli	821
lomake	2
ohje, opas tai suositus	244
opetusmateriaali	1
organisaatiotieto	551
projekti tai kampanja	3
tapahutuma	26
testi	25
tietokanta	1
tietosivu	1361
tilasto	1
tutkimustoiminta	78
usein kysyttyä	2
yhdistystieto	2

Julkaisijan mukaan	
Diabetesliitto	116
Duodecim Lääkärikirja	582
Kansanterveyslaitos	1835
Oulun Omahoito	28
Savonia-ammattikorkeakoulu / Terveystieteiden keskus	137
Terveystieteiden edistämisen keskus	96
Työterveyslaitos	199
UKK-instituutti	276





Miksi TerveSuomi perustuu semanttiseen webiin?



Seitsemän syytä semanttiseen webiin



- ”Älykkäämmät” sovellukset ihmisille
- Yhteentoimivat järjestelmät (interoperability)
- Sisältöjen uusiokäyttö sellaisenaan semantiikan kautta
- Sisältöjen yhdistely ja linkitys (content aggregation)
- Sisällöntuotannon tehostaminen hajauttamalla
- Kielirajojen ylittäminen
- Hajautettu hyödyntäminen verkkopalveluina (mash-up)



Miten tästä eteenpäin?



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Pilottijärjestelmän tuotteistushanke menossa
 - KTL julkaisee lopullisen tuotteen helmikuussa 2009
- Sisällöntuotannon käynnistäminen
 - Laajan kansallisen yhteistyöverkoston kautta



UNIVERSITY OF HELSINKI



SeCo

Palvelun prototyypin julkistustilaisuuden ohjelma



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

13:00	Tilaisuuden avaus	
	Eero Hyvönen Prof., tutkimusjohtaja TKK mediatekniikan laitos ja HY tietojenkäsittelytieteen laitos	
13:05	Kansalliset terveystietohankkeet verkossa	Esittelyt
13:05	Eija Hukka Projektipäällikkö Kansanterveyslaitos	Kansalaisen terveystietoportti TerveSuomi.fi Kansallisen kehityshankkeen esittely http://www.terveysuomi.fi/
13:25	Tuula Kolari Projektikoordinaattori Terveystieteen edistämisen keskus Tekry ry	Järjestöjen terveysportaalihanke TerveSuomi.fi:n rinnakkaishankkeena http://www.tekry.fi/
	TerveSuomi.fi portaali	Visio ja sen toteutus
13:45	Eero Hyvönen Prof., tutkimusjohtaja TKK mediatekniikan laitos ja HY tietojenkäsittelytieteen laitos	TerveSuomi – suomalainen terveystieto semanttisessa webissä http://www.seco.tkk.fi/applications/terveysuomi/
14:10	Kim Viljanen Tutkija TKK mediatekniikan laitos	TerveSuomen semanttinen sisällöntuotanto ja ONKI- ontologiapalvelut http://www.vso.fi/ http://www.seco.tkk.fi/services/onki/
14:30	Osma Suominen Tutkija TKK mediatekniikan laitos	TerveSuomi-portaalin suunnittelu, toteutus ja demonstraatio http://www.seco.tkk.fi/applications/terveysuomi/
15:00	Tilaisuuden päätös	



UNIVERSITY OF HELSINKI



SeCo



HealthFinland - Finnish Health Information on the Semantic Web

Eero Hyvönen, Kim Viljanen, Osmo Suominen
Semantic Computing Research Group (SeCo)
Helsinki University of Technology (TKK) and
University of Helsinki
Finland

<http://www.seco.tkk.fi/>
eero.hyvonen@tkk.fi



Background of HealthFinland



- **FinnONTO – National Semantic Web Ontology Project in Finland [1]**
 - Creating a national infrastructure of ontologies
 - Creating ontology services for utilizing them cost-efficiently
 - » <http://www.seco.tkk.fi/services/onki/>
 - Creating demonstrations of practical applications
 - » “Eating out own dog food”
 - » e.g. MuseumFinland [2] and HealthFinland [3,4]

- [1] Hyvönen et al: **Elements of a National Semantic Web Infrastructure** (invited paper).
Proceedings of ICSC-2007, Irvine, USA, IEEE Press, 2007.
- [2] Hyvönen et al: **MuseumFinland - Finnish Museums on the Semantic Web**.
Journal of Web Semantics, vol. 3, no. 2, pp. 25, 2005
- [3] Hyvönen et al: **HealthFinland - Finnish Health Information on the Semantic Web**.
Proceedings of ISWC-2007, Busan, Korea, Springer-Verlag, 2007.
- [4] Osmo Suominen et al.: **User-centric Faceted Search for Semantic Portals**.
Proceedings of ESWC-2007, Innsbruck, Austria, Springer-Verlag, 2007.



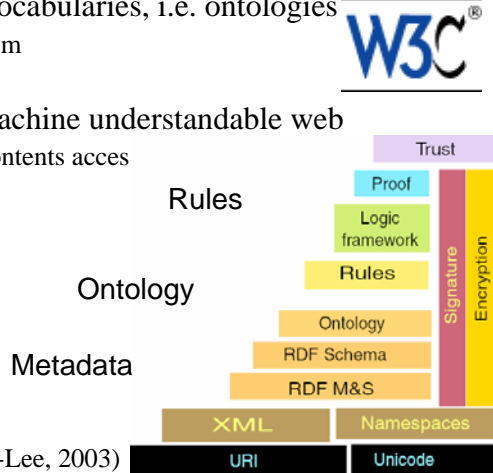
The Semantic Web is Here



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- **Content perspective:** A new metadata layer on the web describing its contents in terms of shared vocabularies, i.e. ontologies
 - Web as a global database system
 - Web of data vs. Web of pages
- **Application perspective:** Machine understandable web
 - The meaning (semantics) of contents acces
 - Enables human usage
 - Intelligent web services
 - Semantic interoperability
- **Technological perspective:** Next layers above XML
 - Generic W3C standards: RDF, OWL etc.
 - Domain specific standards

(Tim Berners-Lee, 2003)



Outline of Talk



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Problems of Mediating Health Information on the Web
 - Citizens perspective
 - Publishers perspective
- Solution approach: semantic web + distributed content creation
 - HealthFinland – Finnish Health Information on the Semantic Web
 - » A semantic portal with distributed content creation
 - Demonstration of the semantic portal
- Conclusions



Problems of Mediating Health Information on the Web



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Citizens view
- Publishers vies



UNIVERSITY OF HELSINKI



SeCo

Problems of Mediating Health Information: **Citizens view**



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Content discovery
 - May need knowledge about publishing organizations
 - Mapping user needs on contents may be difficult
 - » E.g. citizen's vocabulary vs. doctor's vocabulary
- Outdated and missing linkage
- Content aggregation
 - Need content from different perspectives
 - Mismatch between citizens needs and content production
- Quality of content
 - Trust-worthiness
 - Coverage, up-to-date info
- Matching end-user's expertise level
 - Citizens' vs. experts' content



UNIVERSITY OF HELSINKI



SeCo

Problems of Mediating Health Information: Publishers view



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Duplicated work
 - Organizations producing overlapping content
- Difficulty of reusing content
 - Contents created for one purpose/application only
 - Cannot be reused easily in other systems
- Internal and external link maintenance
 - Laborous and expensive
- Indexing (annotation) problems
 - Laborous and expensive
- Quality control
 - Content creation process
 - Content
 - Metadata



HealthFinland Semantic Portal Solution Approach



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Using semantic web technologies
 - For semantic interoperability of content
 - For creating "intelligent" services to citizens
- Using and supporting distributed content creation
- Prototype ready, publication September 12, 2008



Goals of HealthFinland



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Citizen's view
 - **Global view** to health information from different organizations
 - **Aggregated view** to information from different organizations
 - **Intelligent services**
 - » Semantic search
 - Multi-facet search using user-centric views
 - String search based on ontologies
 - » Semantic browsing
 - Semantic recommendations of related information
 - Associative browsing based on ontologies
- Publisher's view
 - 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**



UNIVERSITY OF HELSINKI



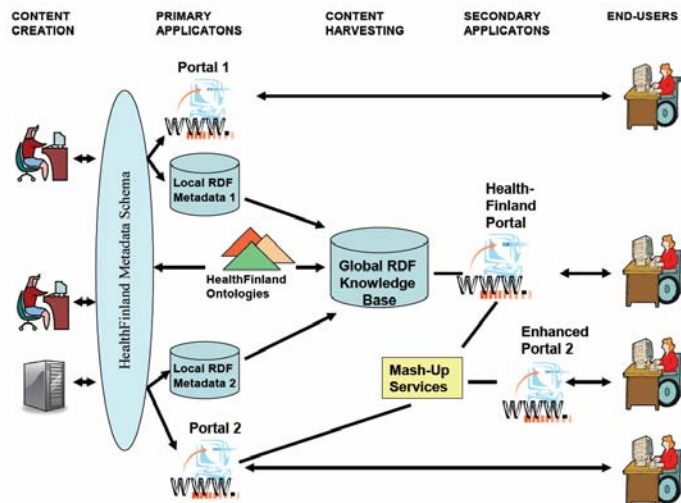
SeCo

HealthFinland

- Finnish Health Information on the Semantic Web System overview



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology



UNIVERSITY OF HELSINKI



SeCo

Technical Solution: Components



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- **Metadata schema** for health information (web pages)
- **Ontologies** for filling in values in the schema
- **Content creation tools**
 - Metadata harvester for collecting and transforming metadata from databases
 - Metadata validator
 - SAHA-editor for annotating content
 - » Can be partly automated using information extraction tool POKA
 - ONKI Ontology servers attached to SAHA using AJAX
- **Semantic portal HealthFinland**
 - Semantic search
 - Semantic browsing



UNIVERSITY OF HELSINKI



SeCo

Metadata Schema in HealthFinland



LOGY

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	Identifier	dc:identifier	1	URI	
	Locator	ts:url	0..1	URL	
	Title	dc:title	1 ^a	Free text	Non-empty string.
	Abstract	dcterms:abstract	1 ^a	Free text	Non-empty string.
	Language	dc:language	1..*	String	RFC 3066
	Publication time	dcterms:issued	1	String	W3CDTF (ISO 8601)
	Acceptance time	dcterms:date:Accepted	0..*	String	W3CDTF (ISO 8601)
	Modification time	dcterms:modified	0..*	String	W3CDTF (ISO 8601)
	Publisher	dc:publisher	1..*	Instance	foaf:Organization
	Creator	dc:creator	0..*	Instance	foaf:Organization, foaf:Person or foaf:Group
Content classification	Subject	dc:subject	1..*	Concept	YSO, MeSH and HPMulti Ontologies
	Audience	dcterms:audience	1..*	Concept	Audience Ontology
	Genre	ts:genre	1..*	Concept	Genre Ontology
	Presentation type	dc:type	1..*	Concept	DCMI Type vocabulary
	Format	dc:format	1	String	IANA MIME types
	Medium	dcterms:medium	1	Concept	Medium Ontology
	Spatial coverage	dcterms:spatial	0..*	String or concept	DCMI Point, DCMI Box or Location 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 Format of	ts:isTranslationOf dcterms:isFormatOf	0..*	Document	URI URI



UNIVERSITY

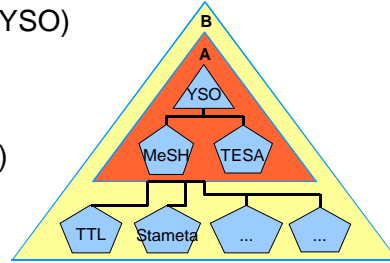
^a Multilingual values are allowed, but only one value in each language.

Ontologies in HealthFinland



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Finnish General Upper Ontology (YSO)
 - >20000 concepts
 - good for everyday terminology
- Medical Subject Headings (MeSH)
 - >20000 concepts
 - good for clinical healthcare (diseases, anatomy etc)
- European Multilingual Health Promotion Thesaurus (HPMULTI)
 - 1200 concepts
 - concepts specific to health promotion
- We had to map these to create one large ontology



UNIVERSITY OF HELSINKI

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



SeCo

25

A Problem of Using Ontologies



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Existing ontologies / thesauri used for annotation
 - a lot of useful structures can be utilized
 - existing metadata uses these anyway
 - interoperability
- **Complex ontologies not suitable for ordinary citizens**
 - 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 intuitive navigation structures for end-users [3]**

[3] Pollit: **The key role of classification and indexing in view-based searching.**
Technical report. University of Huddersfield, UK (1998)



UNIVERSITY OF HELSINKI

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



SeCo

26

Example of the problem



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- 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]



UNIVERSITY OF HELSINKI

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



27

HealthFinland Portal: Faceted Browsing



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Faceted browsing very useful and usable in earlier portals
 - Flamenco [4], SWED [5], MuseumFinland [2] ...
- New idea in HealthFinland
 - Created intuitive facets and categories for end-users
 - End-user interviews, card sorting [6,7]
- Mapped end-user facets to indexing ontologies
 - Separation end-user's and indexer's vocabularies

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

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

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

[7] Suominen, Viljanen, Hyvönen: User-centric faceted search for semantic portals. ESWC-2007.



UNIVERSITY OF HELSINKI

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



28

Facets and Categories



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

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

Secondary facets:

- Publication type
- Publisher
- Publication time
- Target audience

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



UNIVERSITY OF HELSINKI

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



29

Mapping Facets to Ontologies



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Each category contains a set of concepts
- Categories represented using SKOS vocabulary
 - labels: skos:prefLabel
 - hierarchy: skos:broader
- Mappings with SKOS Mapping
 - skosmap:narrowMatch
- 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



UNIVERSITY OF HELSINKI

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

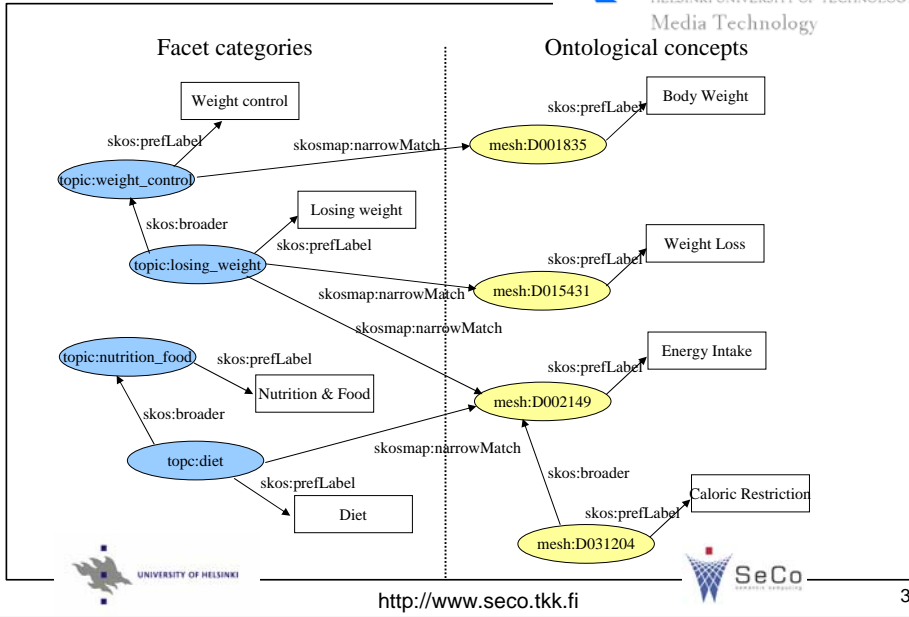


30

Example mappings



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology



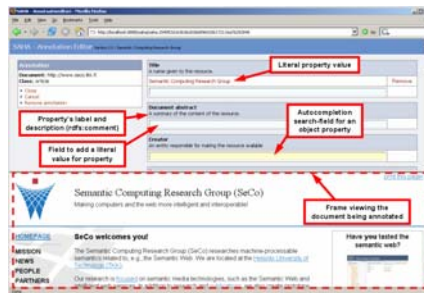
31

Content Creation Tools



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Content harvester
- Metadata validator
- SAHA Annotation editor
- ONKI Ontology Server
 - ISWC-2007 demo session



(Valkeapää, Hyvönen, ISWC 2006 WS)
(Valkeapää, Alm, Hyvönen, JUCS, 2007)



HealthFinland Semantic Portal



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

The screenshot shows the HealthFinland Semantic Portal interface. At the top, there is a search bar and navigation links for 'ETUSIVU' and 'UUTISET'. The main content area is divided into several sections: 'Aihealueet' (Topics) with sub-sections like 'Food and nutrition', 'Exercise', 'Weight control', 'Sexuality', 'Family and children', 'Suurimmat aiheet', 'Intoxicants', 'Catastrophes', 'Epidemics and infectious diseases', 'Mental health and sleep', 'Violence and crises', 'Conditions and symptoms', 'Accidents and first aid', 'Finding help', 'Environment', 'Occupational health', and 'Hakemisto'. A 'Uutiset' (News) section on the right lists recent articles such as 'Kansanterveyslaitosten kansainvälinen järjestö perustettu' and 'Veroale ei lisännyt eläkeikäisten alkoholin käyttöä'. The footer includes the University of Helsinki logo and the URL 'http://www.seco.tkk.fi'.

This screenshot shows the same HealthFinland Semantic Portal interface but with several enhancements. A prominent orange 'Search' box is located at the top right. Below it, a 'Languages' section offers options for 'suomeksi | på svenska | in English'. The 'Uutiset' section on the right is more detailed, showing article titles and dates. A 'Tag Cloud' is visible at the bottom left, and an 'Index' section is highlighted at the bottom center. The overall layout is more structured and user-friendly.

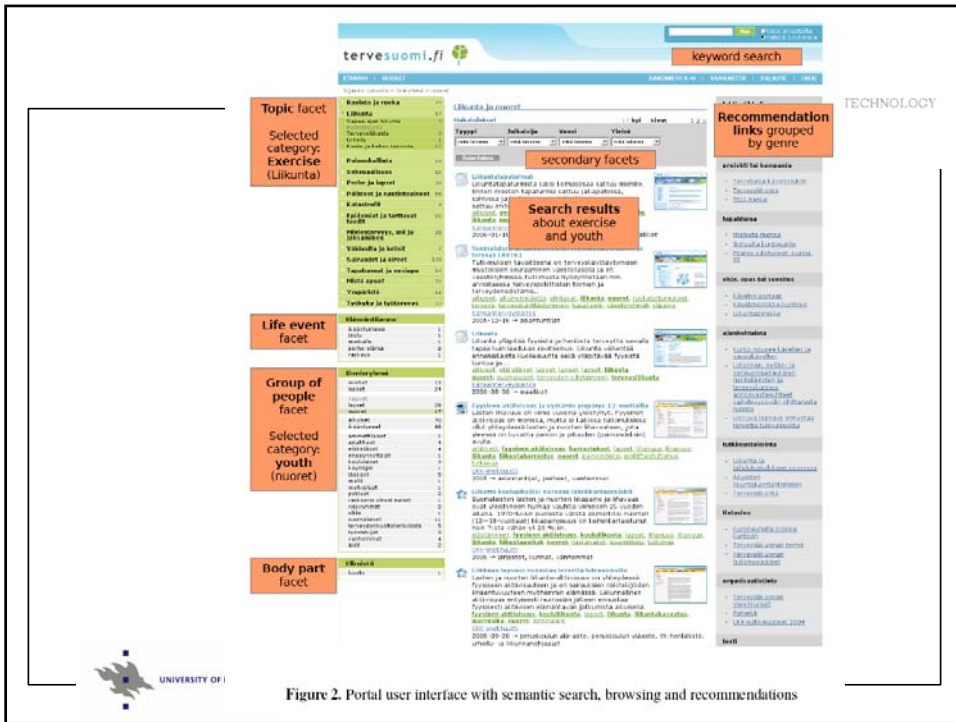
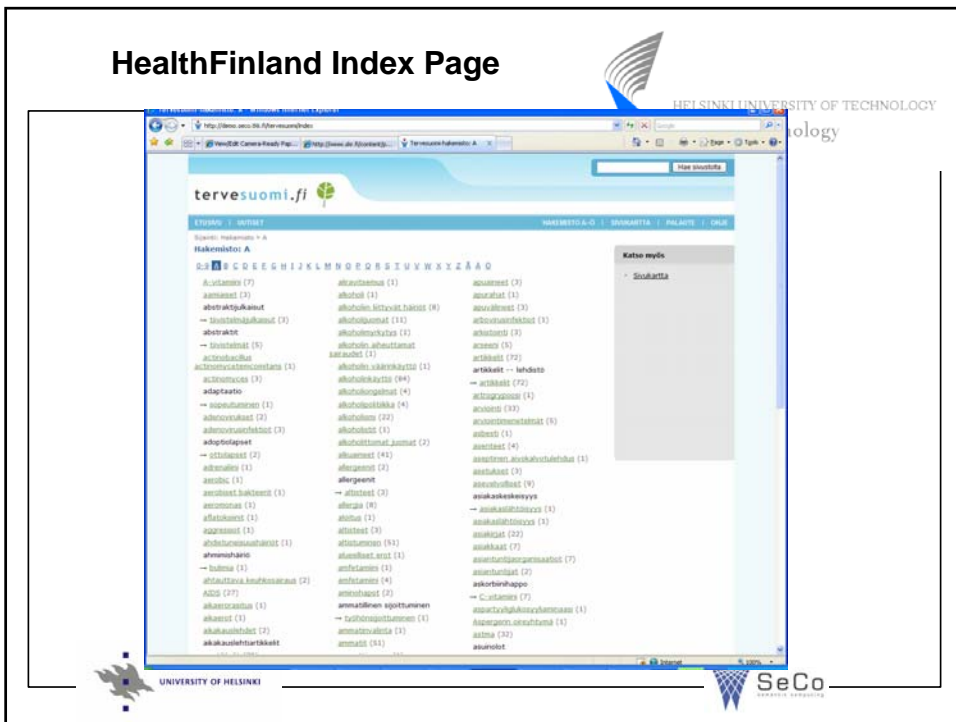


Figure 2. Portal user interface with semantic search, browsing and recommendations



Ontology Browser Index



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

The screenshot shows the 'terveysuomi.fi' website with search results for 'AIDS'. An orange box highlights a search result with the text: "Concept with its sub/super/related concepts". The interface includes a navigation menu on the left, search filters, and a list of search results with various metadata and links.

Conclusions



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- HealthFinland is a **national level publication channel** of health promotion information **on the Semantic Web**
 - Based on shared ontologies and a metadata schema
- Provides citizens with
 - **Global and aggregated view** to heterogeneous information
 - **Semantic search and browsing**
- Provides the publishers with a **collaborative**
 - framework for sharing content production work
 - framework for enriching and linking own content with others
 - tools for making it cost-effective



Publications about HealthFinland



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

2008

Osma Suominen: [Käyttäjakeskeinen moninäkömähaku semanttisessa portaalissa](#). MSc Thesis, February, 2008. [bib](#) [pdf](#)

2007

Eero Hyvönen, Kim Viljanen and Osma Suominen: [HealthFinland - Finnish Health Information on the Semantic Web](#). *Proceedings of the 6th International Semantic Web Conference (ISWC 2007), Busan, Korea*, Springer-Verlag, Nov, 2007. [bib](#) [pdf](#)

Eero Hyvönen, Kim Viljanen, Osma Suominen and Eija Hukka: HealthFinland - Publishing Health Promotion Information on the Semantic Web. *International Journal of Health Care Engineering*, vol. 15, no. 5, pp. 325, oct, 2007. Abstract of a longer paper. [bib](#)

Osma Suominen, Kim Viljanen and Eero Hyvönen: [User-centric Faceted Search for Semantic Portals](#). *Proceedings of the European Semantic Web Conference ESWC 2007, Innsbruck, Austria*, Springer, June 4-5, 2007. [bib](#) [pdf](#)

Osma Suominen, Kim Viljanen and Eero Hyvönen: [Semantic Faceted Search in a Citizens Health Portal](#). Demo track at the European Semantic Web Conference ESWC 2007, Innsbruck, Austria, June 4-5, 2007. [bib](#) [pdf](#)

Osma Suominen, Kim Viljanen, Eero Hyvönen, Markus Holi and Petri Lindgren: [TerveSuomi.fi:n metatietomäärittely \(TerveSuomi.fi metadata specification\)](#). Jan 26, 2007. [bib](#) [pdf](#)

2006

Markus Holi, Petri Lindgren, Osma Suominen, Kim Viljanen and Eero Hyvönen: [TerveSuomi.fi - A Semantic Health Portal for Citizens](#). July 17, 2006. Poster paper, 1st Asian Semantic Web Conference (ASWC2006). [bib](#) [pdf](#)



UNIVERSITY OF HELSINKI

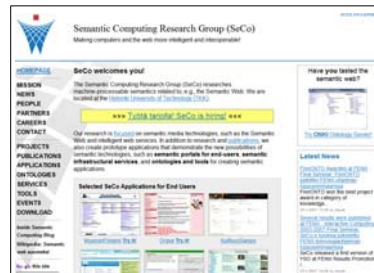
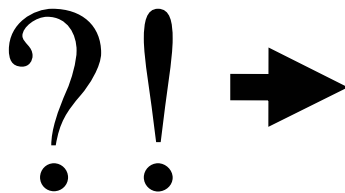


SeCo

Thanks!



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology



www.seco.tkk.fi



UNIVERSITY OF HELSINKI



SeCo

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