

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  1

HELSINKI UNIVERSITY OF TECHNOLOGY  
Media Technology

## Outline of Talk

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



UNIVERSITY OF HELSINKI  2

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


HELSINKI UNIVERSITY OF TECHNOLOGY  
Media Technology

## HealthFinland portal

- | 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


UNIVERSITY OF HELSINKI  3

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

HELSINKI UNIVERSITY OF TECHNOLOGY  
Media Technology

## Goals of HealthFinland: Citizens

- | **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


UNIVERSITY OF HELSINKI  4

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

HELSINKI UNIVERSITY OF TECHNOLOGY  
Media Technology

## Goals of HealthFinland: Publishers

- | 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  5

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

HELSINKI UNIVERSITY OF TECHNOLOGY  
Media Technology

## Challenges

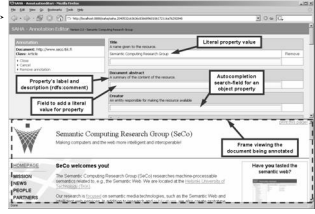
- | create a compelling user experience for the **general public**
  - usability is not just a surface feature!
  - solve actual problems users have – in an intuitive way
- | **tri-lingual portal and KOS**
  - Finnish, Swedish, English
- | **interoperable**
  - with existing KOS
  - legacy metadata, 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**

UNIVERSITY OF HELSINKI  6

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

## Content Creation Tools

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



UNIVERSITY OF HELSINKI <http://www.seco.tkk.fi>

## 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                                  |
|--------------------------|---------------------------|------|-----------------------|--|
| <b>Identifier</b>        | <b>dc:identifier</b>      | 1    | URI                   |  |
| <b>Locale</b>            | <b>dc:locale</b>          | 0..1 | URI                   |  |
| <b>Title</b>             | <b>dc:title</b>           | 1*   | Free text             | Non-empty string                             |
| <b>Abstract</b>          | <b>dc:abstract</b>        | 1*   | Free text             | Non-empty string                             |
| <b>Language</b>          | <b>dc:language</b>        | 1..* | String                | RFC 3066                                     |
| <b>Publication time</b>  | <b>dc:date</b>            | 1    | String                | W3CDTF (ISO 8601)                            |
| <b>Acceptance time</b>   | <b>dc:accepted</b>        | 0..* | String                | W3CDTF (ISO 8601)                            |
| <b>Modification time</b> | <b>dc:modified</b>        | 0..* | String                | W3CDTF (ISO 8601)                            |
| <b>Publisher</b>         | <b>dc:publisher</b>       | 1..* | Instance              | foaf:Organization                            |
| <b>Creator</b>           | <b>dc:creator</b>         | 0..* | Instance              | foaf:Organization, foaf:Person or foaf:Group |
| <b>Subject</b>           | <b>dc:subject</b>         | 1..* | Concept               | YSO, MeSH and HPMulti Ontologies             |
| <b>Audience</b>          | <b>dc:audience</b>        | 1..* | Concept               | Audience Ontology                            |
| <b>Genre</b>             | <b>dc:genre</b>           | 1..* | Concept               | Genre Ontology                               |
| <b>Presentation type</b> | <b>dc:type</b>            | 1..* | Concept               | DCM:Type, vocabulary                         |
| <b>Format</b>            | <b>dc:format</b>          | 1    | String                | IANA, MIME types                             |
| <b>Medium</b>            | <b>dc:medium</b>          | 1    | Concept               | Medium Ontology                              |
| <b>Spatial coverage</b>  | <b>dc:spatial</b>         | 0..* | String or concept     | LSM:Point, LSM:Box or Location Ontology      |
| <b>Temporal coverage</b> | <b>dc:temporal</b>        | 0..* | String or concept     | W3CDTF, DCM:Period or Time Ontology          |
| <b>Part of</b>           | <b>dc:isPartOf</b>        | 0..* | Document              | URI  |
| <b>Rights</b>            | <b>dc:rights</b>          | 0..* | Free text or document | URI or textual description                   |
| <b>Source</b>            | <b>dc:source</b>          | 0..* | Free text or document | URI e.g., ISBN or bibliographical reference  |
| <b>Reference</b>         | <b>dc:references</b>      | 0..* | Free text or document | URI e.g., ISBN or bibliographical reference  |
| <b>Translation of</b>    | <b>dc:isTranslationOf</b> | 0..* | Document              | URI  |
| <b>Form of</b>           | <b>dc:isFormOf</b>        | 0..* | Document              | URI  |

\* Multilingual values are allowed, but only one value in each language.

UNIVERSITY OF HELSINKI <http://www.seco.tkk.fi>

## KOS in HealthFinland

**Source KOS**

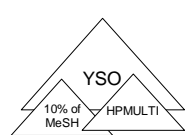
- YSO**  
>20000 concepts  
fi, en, sv  
OWL ontology  
created by FinnONTO (from YSA)  
General knowledge
- MeSH**  
>20000 concepts  
en, fi, sv ...  
SKOS thesaurus  
created by US NLM  
Medical domain (for experts)
- HPMULTI**  
1200 concepts  
en, fi, sv ...  
SKOS thesaurus  
created by EU project  
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>

## User interface and navigation

- 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

UNIVERSITY OF HELSINKI <http://www.seco.tkk.fi>

## 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**

- Anatomy [A]
- Organisms [B]
- Diseases [C]
- Chemicals and Drugs [D]
- Analytical, Diagnostic and Therapeutic Techniques and Equipment [E]
- Psychiatry and Psychology [F]
- Biological Sciences [G]
- Natural Sciences [H]
- Anthropology, Education, Sociology and Social Phenomena [I]
- Technology, Industry, Agriculture [J]
- Humanities [K]
- Information Science [L]
- Named Groups [M]
- Health Care [N]
- Publication Characteristics [V]
- Geographical [Z]

UNIVERSITY OF HELSINKI <http://www.seco.tkk.fi>

## Demonstration of HealthFinland



UNIVERSITY OF HELSINKI <http://www.seco.tkk.fi>

## 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



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



13

## Card sorting



HELSINKI UNIVERSITY OF TECHNOLOGY  
Media Technology

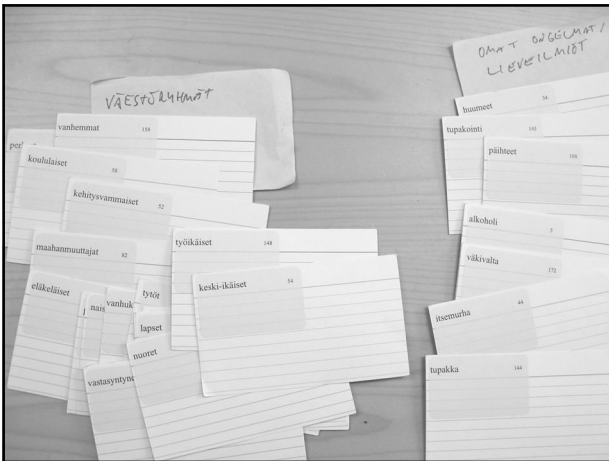
- | method to find out how users conceptualize the information space (i.e. how they group things in their head)
- | often used to build website navigation (information architecture)
  - what should the main sections of the site be
  - how should they be named
  - what things go where
  - what things 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



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



14



## 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

### 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. tooth, nervous system, heart



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



16

## Mapping facets to ontology



HELSINKI UNIVERSITY OF TECHNOLOGY  
Media Technology

- | 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



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

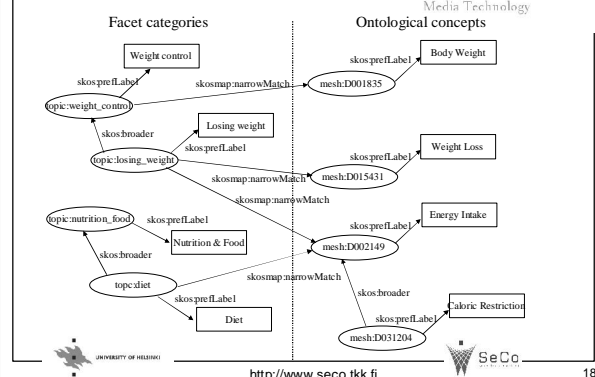


17

## Example mappings



HELSINKI UNIVERSITY OF TECHNOLOGY  
Media Technology



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



18

## Evaluation



- I 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)
- I review of facets by domain experts: some problems corrected
  - too much lumping gives wrong message to users
  - omission of important topics
- I prototype portal to test whether the approach can actually be implemented
- I 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.
- I prototype will go live in early 2008, production use in 2008-2009

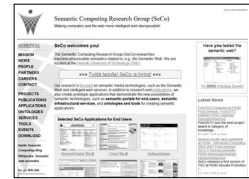
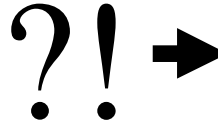


<http://www.seco.tkk.fi>



19

## Thanks!



[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>



20