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/
Outline of Talk

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

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


http://www.seco.tkk.fi
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

http://www.seco.tkk.fi
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**

http://www.seco.tkk.fi
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

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

http://www.seco.tkk.fi
## Metadata Schema in HealthFinland

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

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

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)

http://www.seco.tkk.fi
User interface and navigation

<table>
<thead>
<tr>
<th>Problem: complex ontologies not suitable for human consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>- made for a different purpose (e.g. indexing scientific articles)</td>
</tr>
<tr>
<td>- expert terminology</td>
</tr>
<tr>
<td>- unintuitive hierarchies and groupings</td>
</tr>
<tr>
<td>- everything in a big hierarchy – not facet-based</td>
</tr>
</tbody>
</table>

| need to build navigation structures that users of the site will understand [2] |
| solved using an approach [3] based on card sorting |


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]
Solution approach

- 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


http://www.seco.tkk.fi
Card sorting

- A 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
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

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

Facet categories

- Weight control
- Losing weight
- Nutrition & Food
- Diet

Ontological concepts

- Body Weight
- Weight Loss
- Energy Intake
- Caloric Restriction

http://www.seco.tkk.fi
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 3rd 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.

http://www.seco.tkk.fi
Thanks!

20

osma.suominen@tkk.fi

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

Suominen et al: **User-centric faceted search for semantic portals.**
Proc. ESWC 2007, Innsbruck, Austria, Jun 2007

http://www.seco.tkk.fi