Towards a National Finnish Semantic Web Infrastructure for eCulture

Prof. Eero Hyvönen
Helsinki Univ. of Technology (TKK), Media Technology
University of Helsinki, Dept. of Computer Science
Semantic Computing Research Group (SeCo)
http://www.seco.hut.fi/
Outline of Talk

- The semantic web is coming
- A content infrastructure on the web is needed for it
- What should be done?
  - The vision of a national FinnONTO-project
  - Applications: semantic web at work
A National Problem

- Semantic Web = next generation/layer of the Web
- Ontologies = ”silver bullet” of the Semantic Web
- Finnish ontologies did not exist
- Something should be done about it!

Semantic Web needs a **content infrastructure**

- Like traffic needs roads
- Like energy service needs powerlines, power plants, standards, ...
- Like mobile phones need GSM or 3G-networks
FinnONTO Solution Approach

- Major infrastructure components
  - **Ontologies** to be shared
  - **Ontology services** to utilize ontologies
  - **Standards** to make things interoperable
    » E.g. metadata standards
  - **Tools** to help in creating applications
● Start a national multi-domain ontologization process
  – Making contents of different domain interoperable
  – Thesauri -> ontologies
    » Human usage -> human/machine usage
  – Key ontologies should be open source and maintained publicly
    » Wide acceptance and usage

● Business applications can be built effectively upon a solid infrastructure
Started 2003

2005-2006: 0.8M€ / year
30 funding organizations
(Tekes 80%)

Plan 2006-2007:
38 funding organizations

Memory organizations involved
- National Library of Finland
- Helsinki City Library
- Finnish Literature Society (SKS)
- Espoo City Museum
- National Board of Antiquities
- Finnish Agriculture Museum
- Finnish Museum of Photography
- Ca. 20 city Antikvaria museums
- New members joining in 2006-2007

FinnONTO is lead and is mostly carried out by the Semantic Computing Research Group (SeCo) at the Helsinki Univ. of Tech. and University of Helsinki
Goals

1. **Ontology development** open source
   - General Finnish Ontology based on the national YSA thesaurus (23,000 concepts)
   - Various vertical ontologies based on YSO

2. **ONKI ontology services**
   - Collaborative ontology development
   - Content indexing using web services
   - Ontology-based information retrieval

3. **Pilot applications**
   - Eating our own dog food
1. Ontology development

- Motto: Thesauri -> Ontologies!
Why Thesauri are Not Enough but Ontologies are needed?

- Thesauri with its semantic relations are constructed mainly to help the indexer in her work
  - Understanding the relations needs human knowledge
- The computer could make use of the structure in many application areas:
  - Semantic search and information retrieval
  - Semantic linking of contents
  - Automatic indexing
  - Making contents semantically interoperable
  - ...
- But this is difficult, because the computer is stupid
  - It does not have the human knowledge of an indexer
- Ontologies define accurately the meaning to machines and the humans, too.
Why Thesauri are Not Enough but Ontologies are needed?

- Example from the YSA-thesaurus

- The machine is confused:
  - Is Halley’s comet an individual or a class of them, such as Comet?
  - Can there be many Halley’s comets or only one?
  - Is Comet a kind of Solar system or a part of a solar system. Is it a part as a concept or are all individual comets a part of some solar system?
  - What does ”part of” mean: real part of, contained in, member of, made of, connected to.
  - Do comets have properties of solar systems (e.g. own planets) based on BT
  - Searching ”Solars systems” would retrieve comets although comets are not solar systems
  - ...
Our Practical Solution Approach

- Disambiguate individuals from classes
- Disambiguate major concept meanings
- Refine and disambiguate major meanings in BT, NT, and RT
- Check transitivity of semantic relation chains
- Reorganize and complete the structure into a simple taxonomic ontology
  - Every concept has super concept(s) except the "Thing"

![Diagram]

- Solar system
- Celestial body
- Comet
- Halley's comet

Relations:
- **partOf**: Comet partOf Solar system
- **subClassOf**: Comet subClassOf Celestial body
- **type**: Halley's comet type Comet
A Key Point of Ontologies: Using URIs, not Keywords!

- Each concept will have a **globally unique URI** (across all domains)
  - URI = Universal Resource Identifier
    - URL web addresses are a special case of URIs
    - E.g. "finance:bank", "location:bank", "#GeorgeBush_23", "#Finland", ...
  - A keyword is not enough for indexing the meaning:
    - E.g. "Nokia":
      - = "Nokia" as a company?
      - = "Nokia" as a city in Finland?
      - = "Nokia" a character in a F.E. Sillanpää’s novel?
    - E.g. "Pyhäjärvi" as a location
      - There 49 Pyhäjärvi lakes, villages etc. in Finland
  - The URIs are **globally shared** among users
Idea: Horizontal top ontology + vertical domain ontologies

- Top ontology YSO as a semantic glue
- Merges overlapping domain ontologies
  - Cultural ontology MAO
  - Location ontology SUO
  - Time-location ontology SAPO
  - Actor ontology TOIMO
  - Event & Process ontology TAO
  - Photography ontology VALO
  - Agriforest ontology AFO
- Other classification systems as top ontologies
  » HKLJ + YSO (library domain)
  » ICONCLASS + YSO (fine art domain)
  » MeSH + YSO (medicine domain)
2. Ontology Services & User Groups

1. Ontology Developers
   - Collaborative development of interdependent ontologies
   - Versioning and support for updates

2. Information Searchers
   - Support concept-based search
   - Keyword disambiguation
   - Finding the right search concepts

Nokia: company or city?

2. Information Indexers
   - Support indexing concept finding
   - Keyword disambiguation
   - Support indexing patterns
ONKI-demonstration

- Sharing ontologies on the web when indexing content
- [http://demo.seco.tkk.fi/onki/mao/annotation](http://demo.seco.tkk.fi/onki/mao/annotation)
  - Indexing with concepts (meaning), not with keywords
  - Finding the right indexing annotation concept
  - Retrieving the corresponding URI automatically to an external application
3. Pilot Applications

- eCulture
  - MuseumFinland – Finnish Museums on the Semantic Web
  - CultureSampo – Finnish Culture on the Semantic Web
- eLearning
  - Orava – Semantic video & learning object portal
- eGovernment
  - Semantic Suomi.fi portal
- eHealth
  - Citizens’ health promotion portal Tervesuomi.fi
- Meta-portals
  - Opintoluotsi.fi, Sosiaaliportti.fi, Suomi.fi, ...
MuseumFinland
http://www.museosuomi.fi

- Global seamless view to heterogeneous collections
- Semantic search + browsing
- Common publication channel for museums

(Inter)nationally awarded application
- Semantic Web Challenge Award 2004
- Nordic Digital Excellence in Museums 2004
- Prime Minister’s Innovation Acknowledgement

Google Maps + MuseumFinland
**CultureSampo – Finnish Culture on the Semantic web**

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>KulttuuriSampo</td>
</tr>
<tr>
<td>Artist</td>
<td>Aarre Waldemar Galles-Kallio</td>
</tr>
<tr>
<td>Source</td>
<td>History</td>
</tr>
<tr>
<td>Date</td>
<td>1900</td>
</tr>
<tr>
<td>Medium</td>
<td>Oil on canvas</td>
</tr>
<tr>
<td>Size</td>
<td>102.5 cm</td>
</tr>
<tr>
<td>Material</td>
<td>Linen</td>
</tr>
<tr>
<td>Condition</td>
<td>Good condition</td>
</tr>
</tbody>
</table>

**Description:**
The KulttuuriSampo project is an example of Finnish culture on the Semantic web. It provides detailed information about historical artworks, including artists, dates, and materials, making it easier to access and understand cultural heritage online.
Orava: Video clip & learning object portal

- Semantic search & browsing
  » 2200 videos, Learning Object Metadata (LOM)
- Semantic autocompletion
- Inter-portal linking
  » Linked with MuseumFinland

http://www.museosuomi.fi/orava
Semantic Suomi.fi portal

- Providing alternative views to eGov link library content
- Aggregating relevant content automatically from different organizations
Research Topics Include

- Ontology development
- Semi-automatic annotation
- Ontology mapping and uncertainty
- Semantic search & browsing
- Semantic recommending
- Automatic exhibition construction
- Semantic interoperability
- Semantic meta-search
- Semantic visualization
- Semantic disambiguation
- User interfaces
- Multi-lingual systems
Conclusions

- Semantic web is coming
- An ontology-based infrastructure is needed for it
  - Transforming existing thesauri into ontologies is needed
- Open infrastructure enables development of practical applications
- FinnONTO is an experiment of this on a national Finnish level

- Thank you
- Questions?