Elements of a National Semantic Web Infrastructure - Case Study Finland on the Semantic Web

Eero Hyvönen, Kim Viljanen, Eetu Mäkelä, Tomi Kauppinen, Tuukka Ruotsalo, Onni Valkeapää, Katri Seppälä, Osma Suominen, Olli Alm, Robin Lindroos, Teppo Käänsälä, Riikka Henriksson, Matias Frosterus, Jouni Tuominen, Reetta Sinkkilä and Jussi Kurki

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

- Motivation: Semantic Web coming, how to utilize it in Finland?
- The vision of a national FinnONTO-project 2003-2007
- Goal and results
  - Standards
  - Ontologies
  - Ontology services
  - Tools
  - Applications: semantic web at work
- Conclusions
The Semantic Web Vision

- The problem: The current web designed for humans
  - Natural language, graphics and other multimedia content
- The automatic processing of such content is difficult, because the computer do not have the skills and background information that humans have
- The Semantic Web idea: Encode the information on a conceptually higher level ("semantically") so that even computers can process the information
  - (Based on shared ontologies)
- Goal: Better and more intelligent services for the humans
Perspectives on the Semantic Web

- **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

- **Application perspective:** Machine understandable web
  - The meaning (semantics) of contents accessible to machines
  - Enables
    - Intelligent web services
    - Semantic interoperability

- **Technological perspective:** Next technological layers above XML
  - W3C standards: RDF, OWL etc.
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!

- A solution approach:
FinnONTO Thesis

- 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** for utilizing ontologies
  - **Standards** to make things interoperable
    » E.g. metadata standards
  - **Tools** to help in creating applications
The Way to Go

- 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
FinnONTO project organization

- Universities as research partners
  - Helsinki University of Technology (TKK)
    » Lab. of Media Technology and Lab. of Cartography & Geoinformatics
  - University of Helsinki (UH)
    » Department of Computer Science and Dept. of General Linguistics
  - University of Tampere
    » Department of information sciences
- Directed and 80-90% of the research done by the Semantic Computing Research Group
  - TKK and UH
Industrial & public organization consortium

- 2003-2004
  - 14 funding organizations
- 2004-2005
  - 16 funding organizations
- 2005-2006
  - 30 funding organizations
- 2006-2007:
  - 37 funding organizations
- Current budget 0,8M€ / year (Tekes 80%)
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**
   - Content indexing using ontology web services
   - Ontology-based information retrieval
   - Collaborative ontology development

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

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

- Example from the YSA-thesaurus:

  Solar system
  └── Comet
      ├── Halleys’ comet

  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
  - ...
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
  - 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?
      - = ”Nokia” as an animal?
    - 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 semantic glue
- Merges overlapping domain ontologies
  - Cultural ontology MAO
  - Place ontology SUO
  - Time-place ontology SAPO
  - Actor ontology TOIMO
  - Event & Process ontology TAO
  - Photography ontology VALO
  - Agriforest ontology AFO
  - Ontology of Finnish History HISTO
  - ...

- Aligning other classification systems and top ontologies
  - HKLJ + YSO (library domain)
  - ICONCLASS + YSO (fine arts domain)
  - MeSH + YSO (medical domain)
FinnONTO ontologies

- During the FinnONTO project (2003-2007) many thesauri and other data collections has been transformed to ontologies
  - identifiers (URIs) given to each concept
  - logical consistency checked and fixed
    » e.g., subclassof hierarchies
  - linked to YSO

- in the following, a quick review of 12 most relevant FinnONTO ontologies for describing cultural heritage content
  - (other ontologies of the FinnONTO project are related to, e.g., the health domain)
General Finnish Upper Ontology
YSO

- based on the Finnish General Thesaurus (YSA)
- about 20,000 concepts, e.g., human, children, knowledge, vectors...
- every concept has a URI

- languages: Finnish, (Swedish, English)
  - (Swedish and English needs checking)
- concepts relations in the ontology
  - subclassOf, partOf and associative relation
- can be used as structural model and link to special ontologies
  - different ontologies have common concepts, but include also conflicting information which has to be sorted out
  - has already been merged with the following ontologies or parts of them: Agriforest (agriculture, forestry), MAO (cultural heritage), Design Ontology and Ontology of Photography (Valo), etc.
Cultural Heritage Ontology (MAO)

- concepts for describing cultural heritage content, especially the types of objects
  - e.g., recordings, treasures, car radios, art nouveau...
- contains c. 6700 concepts
- developed originally during the MuseumFinland project
  - is based on the Finnish cultural heritage thesauri MASA, used in most of the Finnish museums
- MAO is currently being mapped to YSO
Design and Communication Ontology (Muoto)

- art and design
  - handicrafts, ceramics, textile design, etc.
- linked to YSO
- related to MAO
- is based on the Finnish Design glossary (c. 2500 terms)
Agriforest ontology (AFO)

- Agriculture and forestry concepts
  - e.g., fish, pine, genes, glucose
- based on the Agriforest thesaurus (c. 6000 concepts)
- linked to YSO
Media Art Ontology

- genres, special concepts, media job descriptions
- e.g. “new media”, “street location art”
Photography Ontology (Valo)

- concepts related to the photography domain
  - techniques, content, etc.
  - e.g., daguerrotype, magnetic photography, form, light
- is based on the thesauri of the Finnish Museum of Photography
- use-case: how to find a photography with a specific mood?
  - a specific exposure, movement, etc
IconClass mapped to YSO

- contains concepts for describing art (paintings etc)
  - e.g., “Adam and Eve in Paradise (before the fall)”
- is mapped to YSO in FinnONTO
- translated to Finnish (by Finnish National Gallery)
Time Ontology (Aika)

- Contains time related concepts
  - e.g., “Middle times”, 19th century, pre-historian time
defines historical events, temporally classified
  – e.g., Declaration of Independence, the Winter war
based on thesauri created by Finnish Agricola researchers
the cultural heritage is most related to historical events
  – different types of items may be related by a certain event
  – historical events are important as content items of their own
Actor Ontology (Toimo)

- Actor Ontology (Persons, Groups and Organizations)
  - Based on Getty Union List of Artist Names (ULAN) [1]
  - Over 100,000 records
  - Mainly artists
- Supports, e.g., relation searching and relation viewing

Akseli Gallen-Kallela, famous painter

related?

Napoleon I, emperor

Example: Actor Ontology (Toimo)
Relations between actors...
Example: Actor Ontology (Toimo)
Relations between actors...
Example: Actor Ontology (Toimo)
Relations between actors...
Finnish Place Ontology (SUO)

- **approximately 150 classes**
  - related to geographical places
    - natural places (e.g. river, lake)
    - man-made places (e.g. building, monument)
  - related to geometric objects
    - i.e. point, curve, polygon and bounding box
- **relationships** between the classes, e.g.
  - mereological relations (2)
  - topological relations (8)
  - geometry valued relations (e.g. isCenterPointOf, 8)
- c. 800 000 **individuals** (instances): Helsinki, Espoo, etc.
- based on Geological Land Survey of Finland's dataset
Finnish Time-Location Ontology (SAPO)

- Based on a method about **ontology time series and coverage-mappings**
- Defines changes and coverages:
  - change bridges: *merged*, *split*, *usedtobe*, *established*, ...
  - ontology mappings: “coverages”: *covers*, *coveredBy*
- 1100 changes and 930 temporal regions are modeled
- An inference engine can reason about the ontology: Lappeenranta (1989-) *covers 12% of Viipuri (-1906).*
- based on Geological Land Survey of Finland's dataset
SAPO Example: Lappeenranta and Viipuri from 1906 until today

Viipuri (-1906) was split in 1906 to Nuijamaa (1906-1944) and to Viipuri (1906-1921)
Metadata standards

- Ontology structures
- Cultural content types
- Health promotion information and services
- Spatial metadata
  - With a national working group
- Learning objects
  - With FinnMeta group
- ...

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

UNIVERSITY OF HELSINKI

SeCo
2. ONKI Ontology Services

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

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

Nokia: company or city?
ONKI Ontology Library Service

- ontologies are (potentially) huge collections of information
- efficient tools are required for
  - ontology development and interlinking ontologies
  - metadata authoring: finding concepts from ontologies
  - end-users: using ontologies for searching
- the ONKI system is FinnONTO's effort towards supporting the users in these tasks
- the current focus in ONKI is on metadata authoring support
  - concept search and browsing
  - adding ontology support (concept search) to legacy systems
  - optimized user-interfaces for ontologies in different domains
- the vision is to build a national ontology service (www.yso.fi)
ONKI for different domains

- the general ONKI service: ONKI-SKOS
  - concept search and browsing, concept fetching
- the geographical ONKI service: ONKI-Geo
  - in addition to ONKI-SKOS functionalities, a map interface for using geographical ontologies
- the actor ONKI
  - extends ONKI-SKOS with optimised visualisations of relations between persons
Demo: ONKI-SKOS
YSO - The Finnish General Upper Ontology

Yso-käsitteet

URI:  http://www.yso.fi/onto/yso/p4205

Context

<table>
<thead>
<tr>
<th>yso-käsitteet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstrakti</td>
</tr>
<tr>
<td>Muuttova</td>
</tr>
<tr>
<td>Pysyva</td>
</tr>
</tbody>
</table>

Properties

<p>| Label (fi):  yso-käsitteet |
| Type:  Concept |</p>
<table>
<thead>
<tr>
<th>Concept</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>ships</td>
<td>Label (en): ships</td>
</tr>
<tr>
<td></td>
<td>Alternative label (fi): matkustajalaivat</td>
</tr>
<tr>
<td></td>
<td>Label (fi): laivat</td>
</tr>
<tr>
<td></td>
<td>Alternative label (sv): passagerafartyg</td>
</tr>
<tr>
<td></td>
<td>Label (sv): fartyg</td>
</tr>
<tr>
<td></td>
<td>Related concept: ferries</td>
</tr>
<tr>
<td></td>
<td>Type: Concept</td>
</tr>
<tr>
<td></td>
<td>Matching concept: laivat</td>
</tr>
<tr>
<td></td>
<td>Superordinate concept: watercrafts</td>
</tr>
</tbody>
</table>

Context:
- yso-kasitteen
- lyysyvä
- lyysinen objekti
- konkreettinen eloton objekti
- conveyances
- watercrafts
- ships

Hide subordinate concepts:
### Museum cataloging system

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material</td>
<td>yes: wool [change]</td>
</tr>
<tr>
<td>Manufacturer</td>
<td>toimo Gallen-Kallela, Aksel (Finnish painter and graphic artist, 1865-1931) [change]</td>
</tr>
<tr>
<td>Place of manufacture</td>
<td>palkkahieski ([sub.ass.])</td>
</tr>
<tr>
<td>Manufacturing time</td>
<td>yso helsinki</td>
</tr>
<tr>
<td>Manufacturing technique</td>
<td>yso</td>
</tr>
<tr>
<td>Used by</td>
<td>toimo</td>
</tr>
<tr>
<td>Place of user</td>
<td>palkka</td>
</tr>
<tr>
<td>Subject</td>
<td>yso</td>
</tr>
<tr>
<td>Measures</td>
<td></td>
</tr>
<tr>
<td>Museum collection</td>
<td></td>
</tr>
<tr>
<td>Responsible museum</td>
<td></td>
</tr>
<tr>
<td>Thesaurus</td>
<td></td>
</tr>
<tr>
<td>Item number</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td></td>
</tr>
</tbody>
</table>

Done
**Museum cataloging system**

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material</td>
<td>yso: wool [change] x</td>
</tr>
<tr>
<td>Manufacturer</td>
<td>toimo: Gallen-Kallela, Akseli (Finnish painter and graphic artist, 1865-1931) [change] X</td>
</tr>
<tr>
<td>Place of manufacture</td>
<td>palkka: Helsinki (tsu: aukkutorit) [change] X</td>
</tr>
<tr>
<td>Manufacturing time</td>
<td>yso: abbreviations [change] x</td>
</tr>
<tr>
<td>Manufacturing technique</td>
<td>yso</td>
</tr>
<tr>
<td>Used by</td>
<td>toimo</td>
</tr>
<tr>
<td>Place of use</td>
<td>palkka</td>
</tr>
<tr>
<td>Subject</td>
<td>yso</td>
</tr>
<tr>
<td>Measures</td>
<td></td>
</tr>
<tr>
<td>Museum collection</td>
<td></td>
</tr>
<tr>
<td>Responsible museum</td>
<td></td>
</tr>
<tr>
<td>Thesaurus</td>
<td></td>
</tr>
<tr>
<td>Item number</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td></td>
</tr>
</tbody>
</table>
Demo: ONKI-Geo
Adding ONKI to legacy systems

- Web 2.0 AJAX/Mash-up web widget concept search component
- can be easily added to any HTML web page with a few lines of JavaScript
Example: Adding ONKI in 10 minutes to a legacy system
Try ONKI Yourself

- http://www.seco.tkk.fi/services/onki/
3. Pilot Applications

- **eCulture**
- **eHealth**
  - Semantic prototype of national citizens’ health promotion portal Tervesuomi.fi (2007)
- **eLearning**
  - Opintie – General learning object portal (2007)
- **eGovernment**
  - Semantic citizen’s Suomi.fi portal (2005)
- **Tools**
- **Other topics**
"MuseumFinland—Finnish Museums on the Semantic Web"

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

- (Inter)nationally awarded application
  - Semantic Web Challenge Award 2004 (2. prize)
  - Nordic Digital Excellence in Museums 2004 (nominee)
  - Prime Minister’s Innovation Acknowledgement

(Hyvönen et al., JWS 2005)

http://www.museosuomi.fi

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

- Generalizing MuseumFinland to cross-domain cultural content
- Modeling cultural context, events, and processes for enhanced intelligence
- Deep event-based semantic annotations
- New interface ideas: semantic autocompetition, maps, time lines, visualizing semantic structures
- New search ideas: searching and explaining links, meta-search
- Successive prototypes

(Junnila et al., FAIS 2006)
(Hyvönen et al., FAIS 2006)
(Kauppinen et al., FAIS 2006)
TerveSuomi.fi ("HealthyFinland.fi")
The citizen’s semantic health promotion portal

- Aggregates content produced by different Finnish health organizations
- Provides information through user-centric views
- Recommends related material of interest

(Holi et al., ASWC 2006)
(Suominen et al., ESWC 2007)
(Hyvönen et al., ISWC 2007)
Orava: Video clip & learning object portal

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

http://www.museosuomi.fi/orava

(Känsälä & Hyvönen, ASWC 2006 WS)
Semantic Suomi.fi portal

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

(Sidoroff & Hyvönen, ISWC 2005 WS)
OntoViews Framework

- Ontogator + Ontodella + Apache Cocoon
- (Makelä, Hyvönen, Viljanen, ISWC 2004)
- (Mäkelä, Hyvönen, Saarela, ISWC 2006)
- (Viljanen; Hyvönen, Känsälä, DEXA WS, 2006)
Semantic Faceted Search Engine
ONTOGATOR

- Integrates view-based search paradigm with semantic web
  - Ontologies and reasoning
- Used to build 8 different systems
- Scales up to millions of search items and hundreds of thousands of categories
- (Mäkelä, Hyvönen, Saarela, ISWC 2006)
Semantic Recommendation Server
ONTODELLA

- For creating semantic recommendations between resources
- Provides
  - logic-based services for projecting facets
  - logic-based recommendations with explanations
- (Viljanen, Hyvönen, Känsälä, DEXA 2006 WS)
Annotation tool SAHA

- Adapts to different annotation schemas
- Support distributed annotation work and ontology population
- Easy to use with a web browser

(Valkeapää, Hyvönen, ISWC 2006 WS)
(Valkeapää, Alm, Hyvönen, JUCS, 2007)
Semi-automatic semantic annotation and authoring tool OPAS

- Support help-desk question-answering (QA)
  - Annotating QA-pairs and helping authoring answers
- National ”Ask the librarian” service as a case study

(Vehviläinen, Hyvönen, Alm, ISWC 2006 WS)
Using and visualizing historical geospatial information

Figure 6: A searchable map interlinked with the semantic portal MuseumFinland.

Figure 4: Using multiple maps simultaneously. A historical Karelian map depicting the park of Monrepos in Viipuri is shown semi-transparently on top of a modern satellite image provided by the Google Maps service.

Figure 5: Search results using the n-point search: Viipuri, Koivisto and Makslahti are matched.

(Kauppinen et al., FAIS 2006)

Figure 3: Annotation and indexing concepts matched.
Other topics

- Uncertainty in ontologies
  - Representing and reasoning with spatial uncertainty
    » (Holi, Hyvönen, Springer 2006)
  - Fuzzy view-based search
    » (Holi, Hyvönen, ASWC 2006)
- Representing and reasoning with historical geospatial changes
  - An ontological model of Finnish communes and counties
    » (Kauppinen, Hyvönen, Springer 2006)
- Semantic autocompletion
  » (Hyvönen, Makelä, ASWC 2006)
- Combining text- and ontology-based searching
- Automatic annotation using NLP
What Next?

- Semantic Web 2.0 – Intelligent Collaborative Services
  - Idea: synergy of Web 2.0 and Semantic Web
  - Based on FinnONTO infrastructure
  - 3 years, 39 funding organizations
  - 2008-2010, if funding application is approved

- And other projects
Conclusions

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

- Papers, software, ontologies, demos, and pilot systems available at: http://www.seco.tkk.fi/

- Thank you
- Questions?
More info: SeCo publications online 2007
http://www.seco.tkk.fi/publications/