



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!
-
- A solution approach:
The National Finnish Semantic Web
Ontology Project FinnONTO (2003-2007)



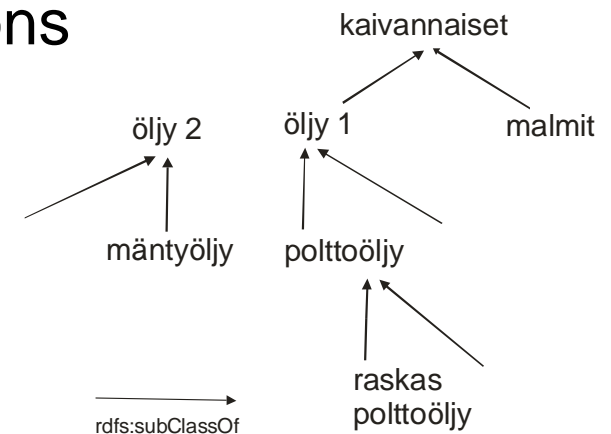
- Semantic Web needs a **content infrastructure**
 - Like traffic needs roads
 - Like energy service needs powerlines, power plants, standards, ...
 - Like mobile phones need GSM or 3G-networks

FinnONTO Solution Approach



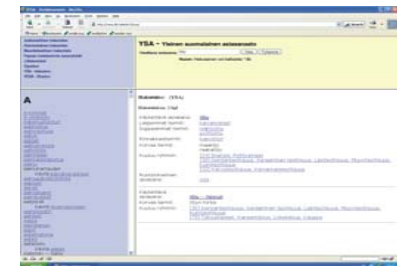
HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- 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



Partnering organizations

Companies and Organizations

- [AAC Global Oy](#)
- Antikvaria-ryhmä
- [AlmaMedia Oyj](#), 2003-2005
- [Blue Meteorite Ltd](#) (Sininen meteoriitti Oy)
- [Connexor Oy](#)
- [Elisa Oyj](#)
- Espoo City Museum (Espoon kaupunginmuseo)
- [Finnish Agriculture Museum](#) (Suomen maatalousmuseo)
- [Finnish Broadcasting Company YLE](#) (Yleisradio Oy)
- [The Finnish Centre for Technical Terminology](#) (Tekniikan Sanastokeskus ry)
- [Finnish Literature Society](#) (Suomalaisen kirjallisuuden seura)
- [Finnish National Board of Education](#) (Opetushallitus)
- [The Finnish National Gallery](#) (Valtion taidemuseo)
- [The Finnish Museum of Photography](#) (Suomen Valokuvataiteen museo)
- [The Finnish Terminology Centre](#) (Sanastokeskus TSK)
- [Geological Survey of Finland GTK](#) (Geologian tutkimuskeskus GTK)
- [Grip Studios Interactive Oy](#)
- [Helsinki City Library](#) (Helsingin kaupunginkirjasto)
- [Ministry of Agriculture and Forestry](#) (Maa- ja metsätalousministeriö)
- [Ministry of Finance](#) (Valtiovarainministeriö)
- [The National Board of Antiquities](#) (Museovirasto)
- [National Land Survey of Finland](#) (Maanmittauslaitos)
- [The National Library of Finland](#) (Kansalliskirjasto)
- [National Public Health Institute KTL](#) (Kansanterveyslaitos)
- [National Research and Development Centre for Welfare and Health \(STAKES\) / Luokituskeskus](#)
- [Tekes, the Finnish Funding Agency for Technology and Innovation](#)
- [M-Cult ry](#)
- Opintoluotsi-project/OPM
- [Sosiaaliportti-project / STAKES](#)
- Suomen maatalousmuseosäätiö
- [TietoEnator Corporation](#)

Research organizations

- [Helsinki University of Technology, Institute of Cartography and Geoinformatics](#) (prof. Kirsi Virrantaus, subproject director)
- [Helsinki University of Technology, Laboratory of Media Technology](#) and [University of Helsinki, Department of Computer Science](#), (prof. Eero Hyvönen, project director)
- [University of Helsinki, Department of General Linguistics](#) (prof. Lauri Carlson, subproject director)
- [University of Tampere, Department of Information Studies](#) (prof. Kalervo Järvelin, subproject director)

Started 2003

2005-2006: 0,8M€ / year
30 funding organizations
(Tekes 80%)

Plan 2006-2007:
38 funding organizations



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

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 the Semantic Computing Research Group (SeCo) at the Helsinki Univ. of Tech. and University of Helsinki



UNIVERSITY OF HELSINKI



Goals



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- **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!

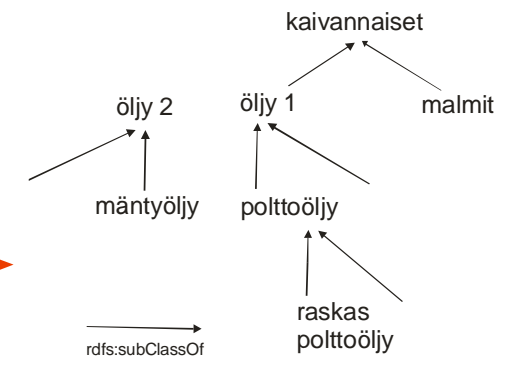
YSA

The screenshot shows the YSA website interface. The search results for 'öljy' are as follows:

- Hakutulos: (YSA)**
- Hakunäköala: 2 kpl
- Käytettävä asiasana: [öljy](#)
- Laajemmät termit: [kaivannaisto](#)
- Suppeammat termit: [mäntyöljy](#), [polttoöljy](#)
- Rinnokkaistermit: [kasviöljy](#), [maaoöljy](#), [raakaöljy](#)
- Kuuluu ryhmiin: [\[19\] Energia_Polttoaineet](#), [\[30\] Kemianteollisuus_Keraaminen teollisuus_Lasiteollisuus_Muoviteollisuus_Kumiteollisuus](#), [\[32\] Kaivosteollisuus_Kaivannaisteollisuus](#)
- Ruotsinkielinen asiasana: [olja](#)
- Käytettävä asiasana: [öljy -- hinnat](#)
- Korvaa termit: [öljyn hinta](#)
- Kuuluu ryhmiin: [\[30\] Kemianteollisuus_Keraaminen teollisuus_Lasiteollisuus_Muoviteollisuus_Kumiteollisuus](#), [\[75\] Taloustieteet_Kansantalous_Liiketalous_Kauppa](#)

Esimerkkiongema:
kaivannaiset ST: öljy ST: mäntyöljy,
mutta mäntyöljy ei kuulu kaivannaisiin!

YSO



Why Thesauri are Not Enough but Ontologies are needed?



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

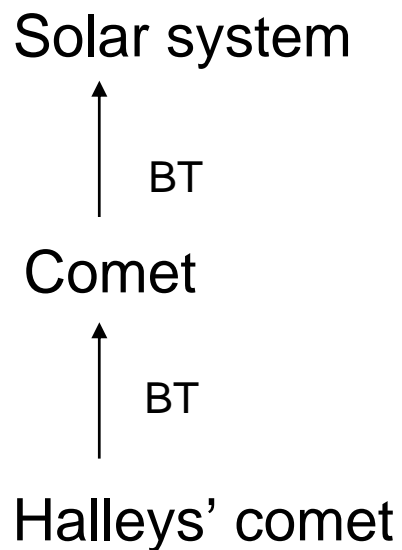
- Thesauri with its semantic relations are constructed mainly to help the indexer in her work
 - Understanding the relations needs human knowledge
- The computer could make use of the structure in many application areas:
 - Semantic search and information retrieval
 - Semantic linking of contents
 - 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?



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Example from the YSA-thesaurus



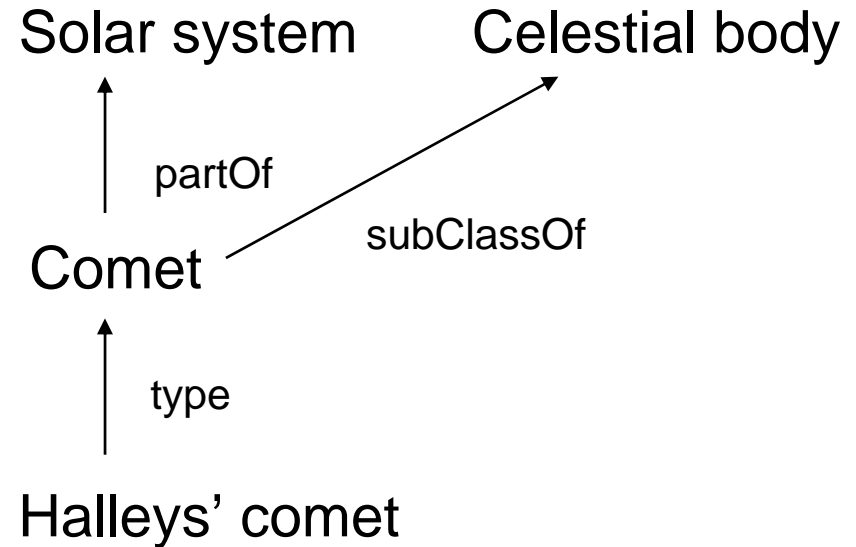
- The machine is confused:

- 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"



A Key Point of Ontologies: Using URIs, not Keywords!



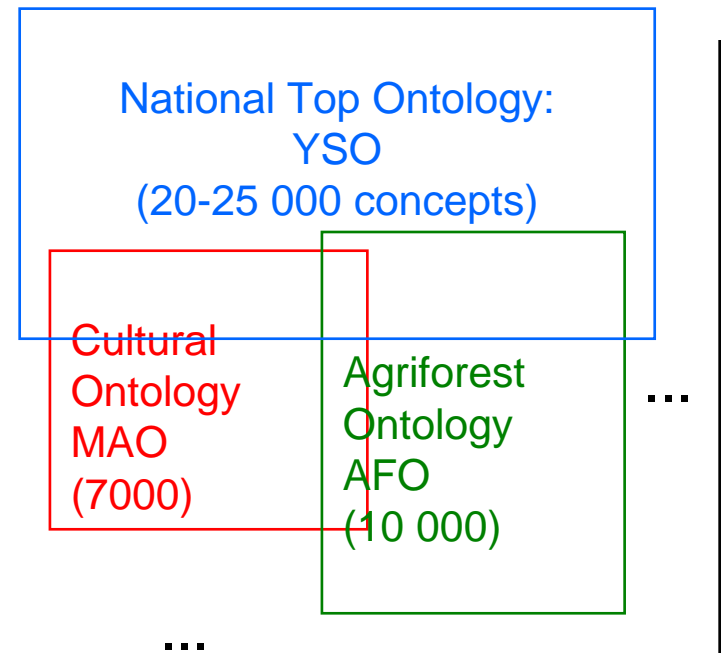
- Each concept will have a **globally unique URI** (across all domains)
 - URI = Universal Resource Identifier
 - » URL web addresses are a special case of URIs
 - E.g. "finance:bank", "location:bank", "#GeorgeBush_23", "#Finland", ...
 - A keyword is not enough for indexing the meaning:
 - » E.g. "Nokia":
 - = "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



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- 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

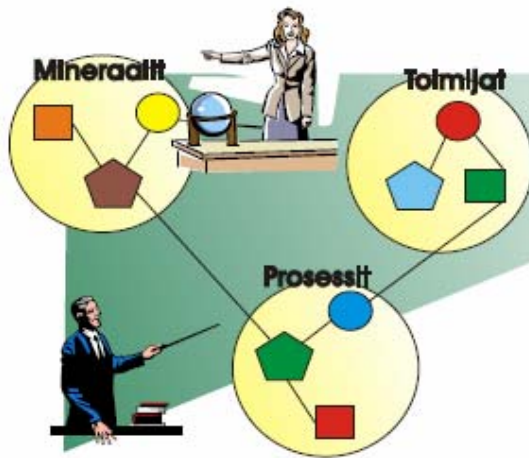
- Collaborative development of interdependent ontologies
- Versioning and support for updates

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>
 - 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, ...





Google Maps + MuseumFinland

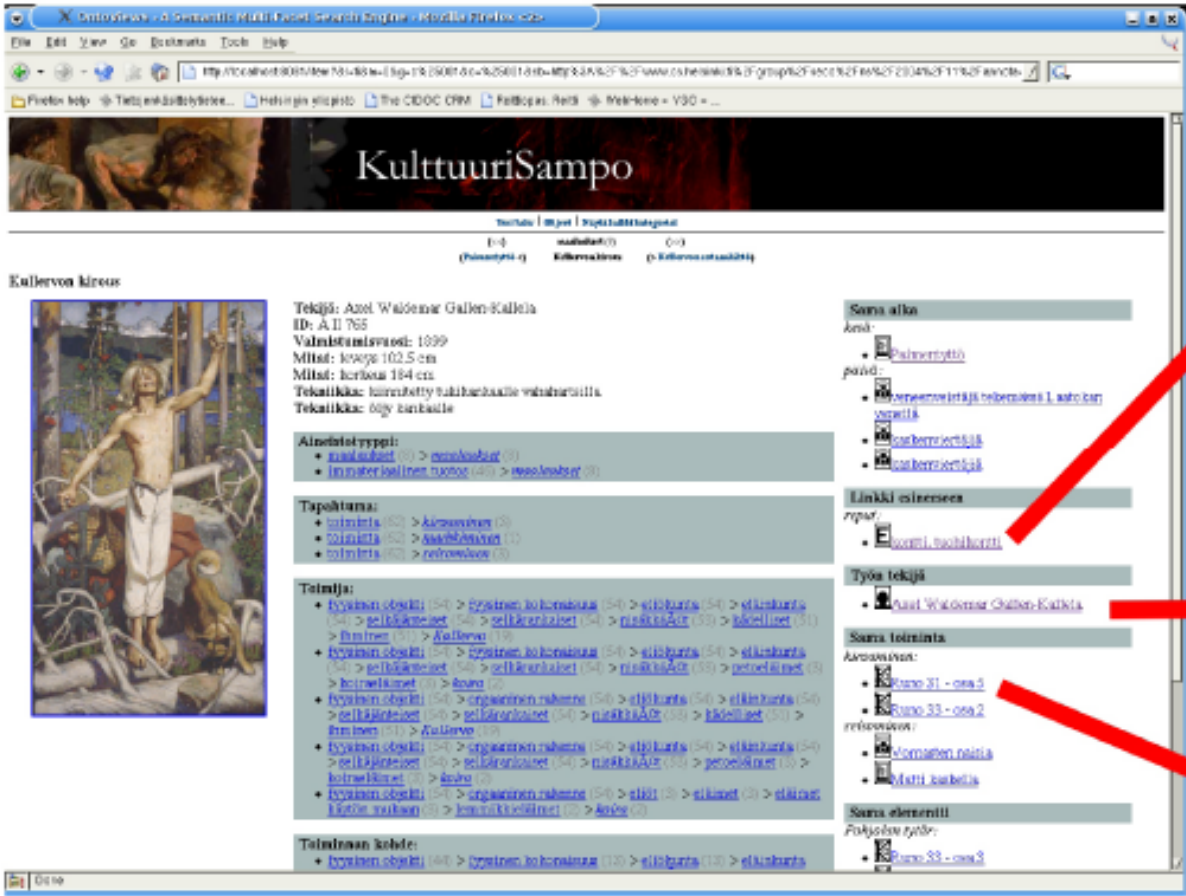
- Ideas
 - Global seamless view to heterogeneous collections
 - Semantic search + browsing
 - Common publication channel for museums
- (Inter)nationally awarded application
 - Semantic Web Challenge Award 2004
 - Nordic Digital Excellence in Museums 2004
 - Prime Minister's Innovation Acknowledgement



CultureSampo – Finnish Culture on the Semantic web



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology



Kullervo ja Kyy
Kullervo lauletti kun hän tahtoikin,
Kullervo ja Kyy kullervon kireän laulun.
Puno 31 - osa 5
En tiedä, kuka painoi,
kalle työlle työstönsä.
Punon kullervo kullervon
Punon kullervo kullervon
Kullervo, Kullervon poika,
taasen taas ainoaksi riikiksi:
"Kullervo, kullervo kullervo,
kullervo kullervo kullervo,
kullervo kullervo kullervo,
kullervo kullervo kullervo."
Kullervo kullervo kullervo.

Orava: Video clip & learning object portal



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

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

<http://www.museosuomi.fi/orava>

Orava: Opetusvideoiden haku- ja suositteleva

Opetusvideoiden haku- ja suositteleva

Tervetuloa Orava-portaaliin!

Orava tarjoaa helpon tavan selata Yle:n Klaffi-portaalin aineistoa yksinkertaisen käyttöliittymänsä kautta.

Aloita valitsemalla sinua kiinnostava kategoria vasemmassa laidassa sijaitsevista vaihtoehdoista, jolloin sivun keskiosaan tulee näkyviin linkkejä videoiden kohdesivuille.

Voit tarkentaa hakua valitsemalla vasemmasta laidasta lisää kategorioita tai tutkia jotakin mielenkiintoista videota lähemmin napsauttamalla hakutuloksesta videon linkkiä.

Hakuehdon polistaminen tapahtuu napsauttamalla uudelleen valittua kategoriaa, joka on esitetty korostettuna.

Tarkempia ohjeita löytyy oikean yläkulman ohje-linkkiä napsauttamalla.

Portaalin on tehnyt Helsingin yliopiston tietojenkäsittelytieteen laitoksen ohjelmistotuotantoprojekti-kursseille ryhmä Orava

Suosituimmat kohteet

- Kuva mielessä: katseen kulku
- Supisuomea: Helsinki on Suomen pääkaupunki

Viimeksi katsotut kohteet

- Kuva mielessä: katseen kulku
- Supisuomea: Helsinki on Suomen pääkaupunki
- Kansalaisen ABC: Tietoyhteiskunta, uusi työelämä, sql-kieli ohjelmoinnin pohjana
- Televisio: henkilö, Esa-Pekka Tiitinen
- Opinkino: pääjuttu, Opetushallitus



UNIVERSITY OF HELSINKI



SeCo
SEMANTIC COMPUTING

Semantic Suomi.fi portal



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

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

SW-Suomi.fi
- A Semantic Information Portal -

Suomeksi | New search | Instructions (in Finnish) | Show all categories

Concept search: Find

Topic (group hits)(category tree)	Audience (group hits)(category tree)	Regional (group hits)(category tree)
Home and family (35), Education and Libraries (30), Culture and Recreation (2), Transport and travel (10), Immigration and emigration (4), Rights and security (7), Money and property (3), Health and nutrition (10), Income security and pensions (1), Society and Citizens (9), Environment and nature (2)	Pensioners (2), Households (24), Kids (11), Young people (20), Students (12), Families (19), Persons seeking employment (3), Employees (3), Finns living abroad (1), Minorities (2), Disabled persons (1), Elderly (1), Offices and bureaus (1), Citizens (84), Enterprises (6)	paikat/finland (118), paikat/ulkomaat (16) Language (group hits)(category tree) English (2), Finnish (150)
By content (group hits)(category tree) Contact information (14), Electronic services (4), Legal forms (6), Discussion forums (4), Information and guides (49), Legal information (14), News (2), Statistics (3), Study materials (1)	Life event (group hits)(category tree) New home (8), Moving (10), Starting studies (11), Family life (25), Falling ill (5), Work and career (9)	



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



HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- 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?

