Semantic Web Technologies in Content Management

National Finnish Ontology Project (FinnONTO)

Tomi Kauppinen, M.Sc., Researcher

tomi.kauppinen [at] tkk.fi

Semantic Computing Research Group
Laboratory of Media Technology
Helsinki University of Technology

Department of Computer Science
University of Helsinki

Project Consortium (2005-06-08)

Helsinki University Library
THE NATIONAL LIBRARY OF FINLAND

TEKES

National Board of Antiquities

KIASMA

ALMA | MEDIA

The Finnish Museum of Photography

AAC GLOBAL

TietoEnator
Building the Information Society

m-cult

SanastokeskusTSK

Connexor Oy

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

UNIVERSITY OF HELSINKI

UNIVERSITY OF TAMPERE

Finnish National Ga

SeCo
semantic computing

UNIVERSITY OF HELSINKI

HELSINKI UNIVERSITY OF TECHNOLOGY

Semantic Web: Technology push

Trust level
Digital signature, annotations,...

Logic level
KIF, RuleML, SWRL, ...

Ontology level
RDFS, OWL, WordNet, ...

Metadata level
RDF, Topic Maps,...

Structure level
XML, XML DTD/ Schema, XSL,...

Internet level
Unicode, URI,...

Planning
CPR, SPAR, PDDL,…

Processes
BPML, WPDL, PSL,…

Services
UDDI, WSDL, OWL-S,…

Transactions
XML/EDI, KQML,…

Communication
TCP/IP, HTTP, SOAP,...

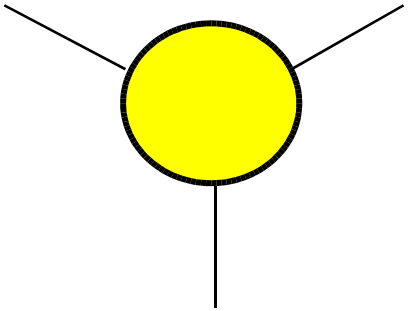
What is new in Semantic Web?

PROGRAMMING

Object-oriented modeling

ARTIFICIAL INTELLIGENCE

Knowledge representation
Logic based semantics



XML-syntax, e.g., RDF(S)

WWW-TECHNOLOGIES

National Ontology Project in Finland

- **Establishing the semantic infrastructure for the Finnish semantic web**
  - **Yleinen Suomalainen Ontologia (YSO)**
- Motivations for the project
  - Thesauri and classification systems have been widely used for indexing contents
- There are, however, many problems involved!
  - Current thesaurus-based approach is not good enough for the semantic web
  - Machine understandability missing
- Funded by Tekes and a large consortium of companies and institutions
  - 9/2003-8/2005

Ontologies

- Facts:
  - Ontologies = The Core of the Semantic Web
  - Semantic Web = The next generations of WWW
  - There are NO Finnish ontologies available!
- Theses underlying the Project
  - Core ontologies should be open source
    - This maximizes **usage and interoperability**
    - This maximizes **business opportunities**
  - Core ontologies should be created together at the ”national” level
    - Wide commitment is needed (”standards”)
    - Maintenance by public institutions needed
  - **This is a question of the national IT infrastructure**

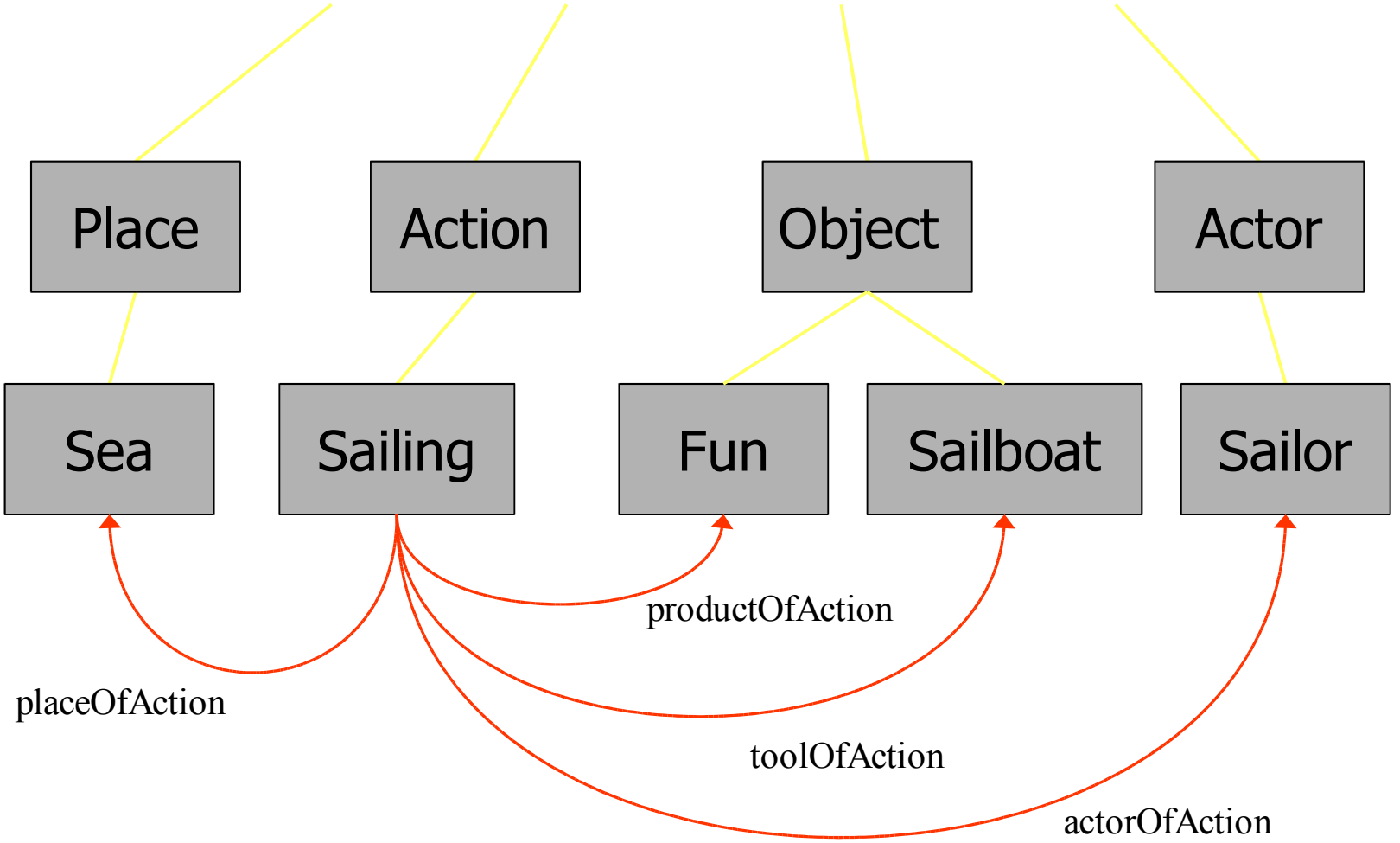
Ontologies

- "An ontology is an explicit specification of a conceptualization" (Gruber, 1995)
- "A conceptualization is a set of conceptual relations defined on a domain space" (Guarino, 1998)
- Expressed in a formal machine understandable way (e.g. RDF(S), OWL)
- Examples: WordNet, Standard Upper Ontology (SUO), dmoz.org, TAP, MAO, FRBR, CIDOC CRM, DOLCE, …
- Define vocabulary for metadata formats
  - e.g. Dublin Core, LOM, …

Ontologies produced in the project

- A top ontology **YSO (Yleinen Suomalainen Ontologia)**
  - **Over 28 000 concepts** ranging e.g. from banking and geography to arts and mathematics
  - based on the general Finnish keyword thesaurus YSA
    - http://vesa.lib.helsinki.fi
  - Terms in Finnish
    - bilingual (Swedish)
- Domain ontologies related to YSO
  - **Culturico-historical**
    - MAO, OCM (Outline of Cultural Material)
  - **Art**
    - ICONCLASS, mediaculture, photography
  - **Locations** (at different times), SAPO
  - **Actors** (persons, organizations, etc.), SUTO

Example 1: an ontology



Why use ontologies in content description?

1. Ontologies overcome problems related to thesauri
2. Use of ontologies has advantages for content description

Problems of thesauri

- Interoperability
- Identification of concepts
- Semantics too simple
- Managing large thesauri
- Managing changes

1. Interoperability

- Systems are hetegoneous
  - Syntax: systems use different data syntax/models
  - Semantics: Terminology of different application fields, organizations and catalogers differ
- The result: Systems cannot operate together

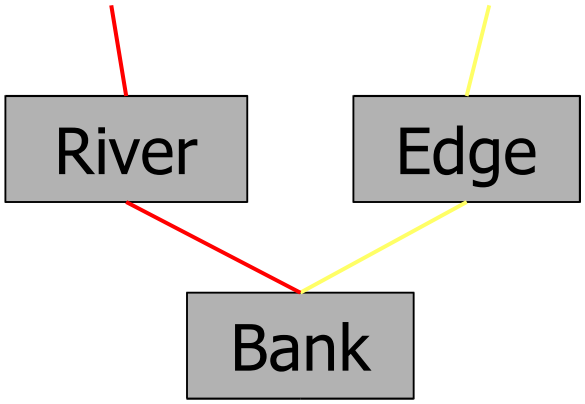
- Ontologies are expressed in a formal **machine understandable** way (RDF(S), OWL)
- Formal languages can be **automatically translated** into other formal languages using existing tools

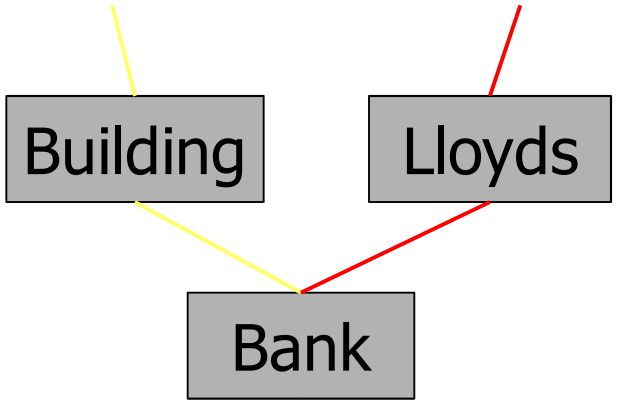
2. Identification of concepts

- In thesauri homonymy need to be distinguished
- Polysemy is hard to disambiguate

- Homonymic words can be **distinguished with URI** (Uniform Resource Identifier)
- Corresponding concepts can be **disambiguated by** their **context**
- Can be applied to meronymic terms as well

Example: Handling homonymies





RiverBank

BankBuilding

3. Semantics too simple

- The semantic system (NT, BT, RT, etc.) is too simple for creating truly intelligent systems
  - Dealing with uncertainty and fuzzy concepts
  - E.g. meronymies, different associative relations etc. are needed

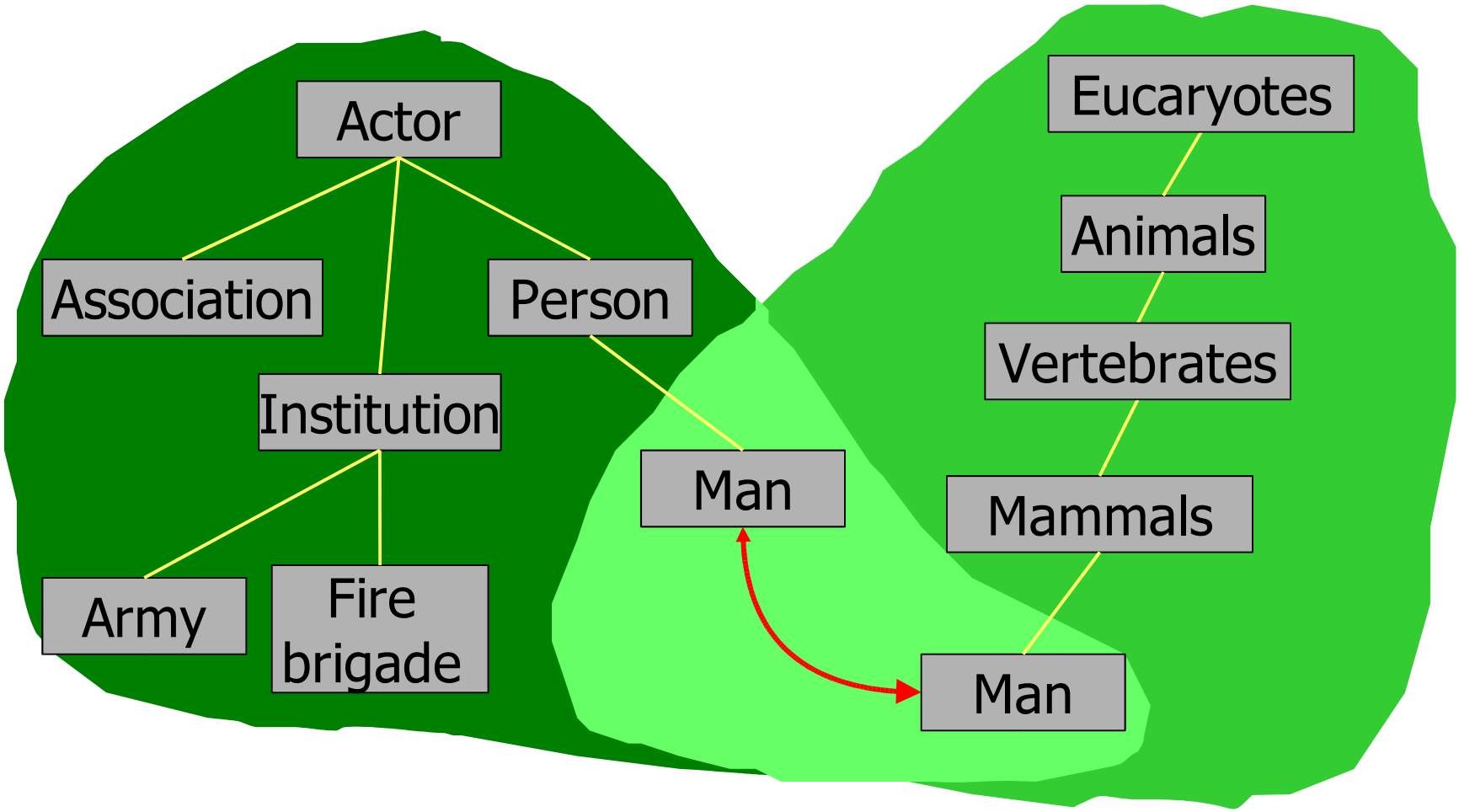
- Ontologies provide a way to define **new relations** when needed.
- Relations can also be defined in a more **detailed** manner: name, domain, range, symmetric/transitive etc.
- Ontologies make it possible to also have **grouping concepts** that do not mix with the ”real” terms.

4. Managing large thesauri

- No organization is capable of maintaining the thesauri of all fields
- The work has to distributed to different expert groups working together

- Distributed maintenance of ontologies can be **assisted by computers**
- Ontologies can also be **mapped with other ontologies:** reaching a large web of **interoperable semantics** where all the individual ontologies are maintained by the experts of that area
- **Open source** would boost application development

Example: Managing large ontologies



Example: Ontologies in MuseumFinland

Seven main ontologies of MuseumFinland:

ONTOLOGY	CONTENT	CLASSES	INDIVIDUALS
Artifacts	Classes for tangible collection objects	3227	0
Materials	Substances that the artifacts are made of	364	0
Actors	Persons, companies, organizations, and other agents	26	1715
Locations	Continents, countries, cities, villages, farms etc.	33	864
Times	Eras, centuries, etc. as time intervals	57	0
Events	Situations, events, and processes in the society	992	0
Collections	Museum collections included in the system	22	24

Altogether about 10,000 interrelated concepts

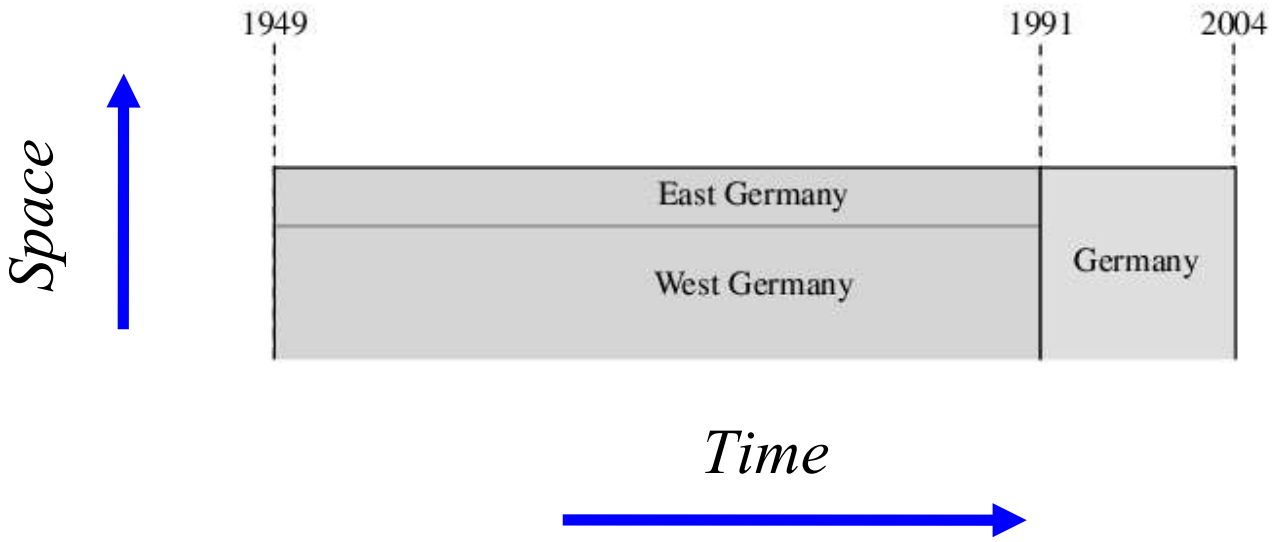
5. Managing changes and time

- The thesauri and concepts change over time
- New concepts emerge all the time, Czechoslovakia does not exist any more, Petsamo is not part of Finland today, etc.
- The contents are indexed with old keywords/concepts but may be retrieved with new ones

- **Ontology versions** can be made interoperable by **bridging** them: annotated objects can be found through both the old and the new concepts
- **Validity time** can be attached to the concepts and logics can be used to reason about them
- **No confusion** about changes in time, e.g. Petsamo of today is not optimal for annotating object dating to 1930's

Suomen Ajallinen PaikkaOntologia (SAPO)

- **Problem: Due to changes in geographic regions, annotation of items in museums and libraries is hard.**
  - **An example: East Germany and West Germany were *merged 1991 to form Germany*.**

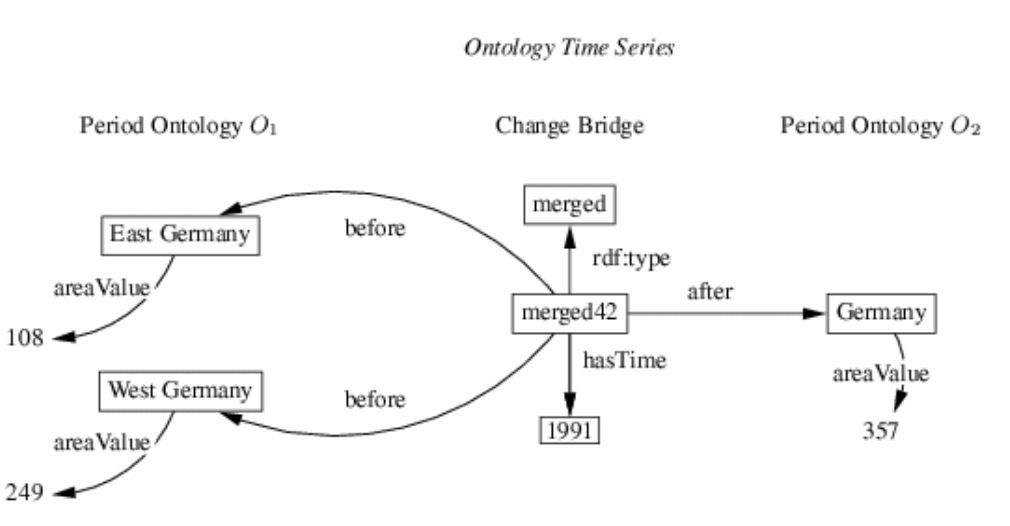


Features of the *OntoFlux*-method

- Changes are defined as an ontology.
- Each region has an own identifying URI.
- Changes are bridged using specific change mappings, “change bridges”:
  - *merged, split, usedtobe, ...*
- Change bridges are transformed automatically to a local coverage graph and then to a global coverage graph.
- An inference engine reasons about local and global coverages between regions.
  - Lappeenranta (1989-) *covers 12% of Viipuri (-1906).*

Suomen Ajallinen PaikkaOntologia (SAPO)

- **Solution: Changes in regions are bridged using Semantic Web -technologies and modeled as an ontology time series**
  - **An inference engine reasons about coverages between the temporal regions of the ontology**



Suomen Ajallinen PaikkaOntologia (SAPO)

- **From the beginning of 20th Century, there are over 1100 changes (merges, splits, name changes) in Finnish counties.**
  - **An Example: *Nuijamaa itsenäistyi Viipurista 1906. Nuijamaa liitettiin Lappeenrantaan 1989***
  - ***Changes are collected by Geological Survey of Finland.***

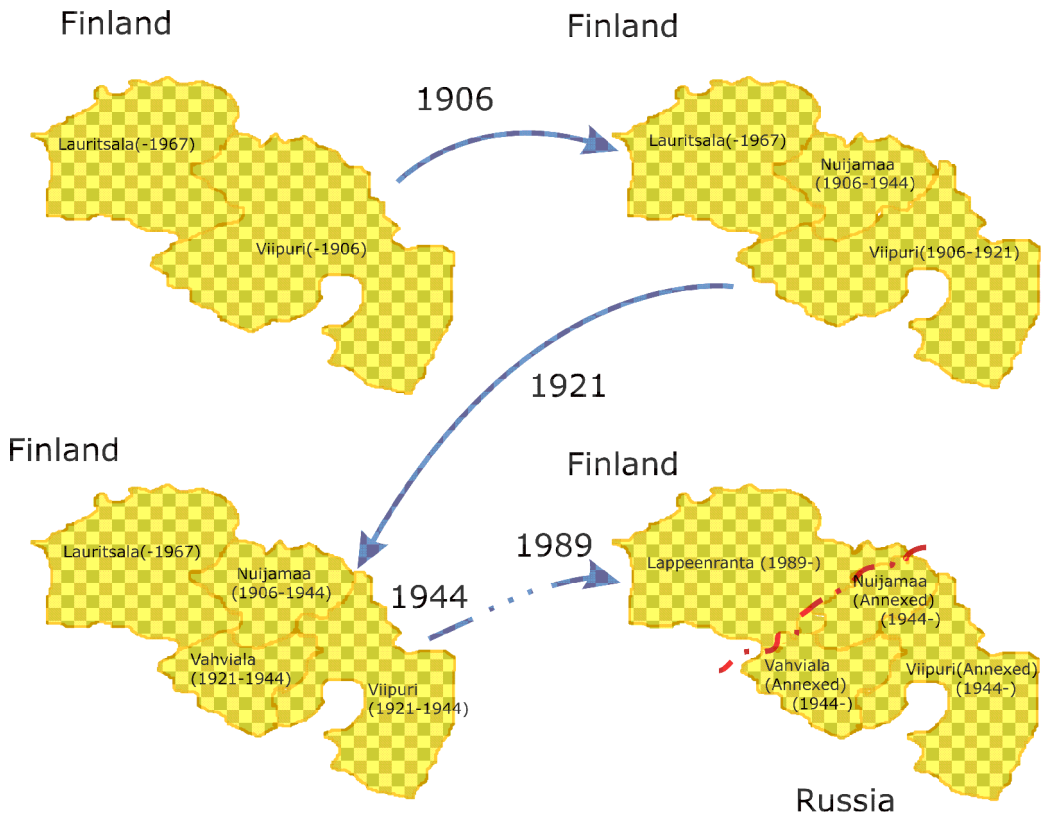
An example of changes in Finnish regions

ESWC 2005, Greece
Best Poster Award!

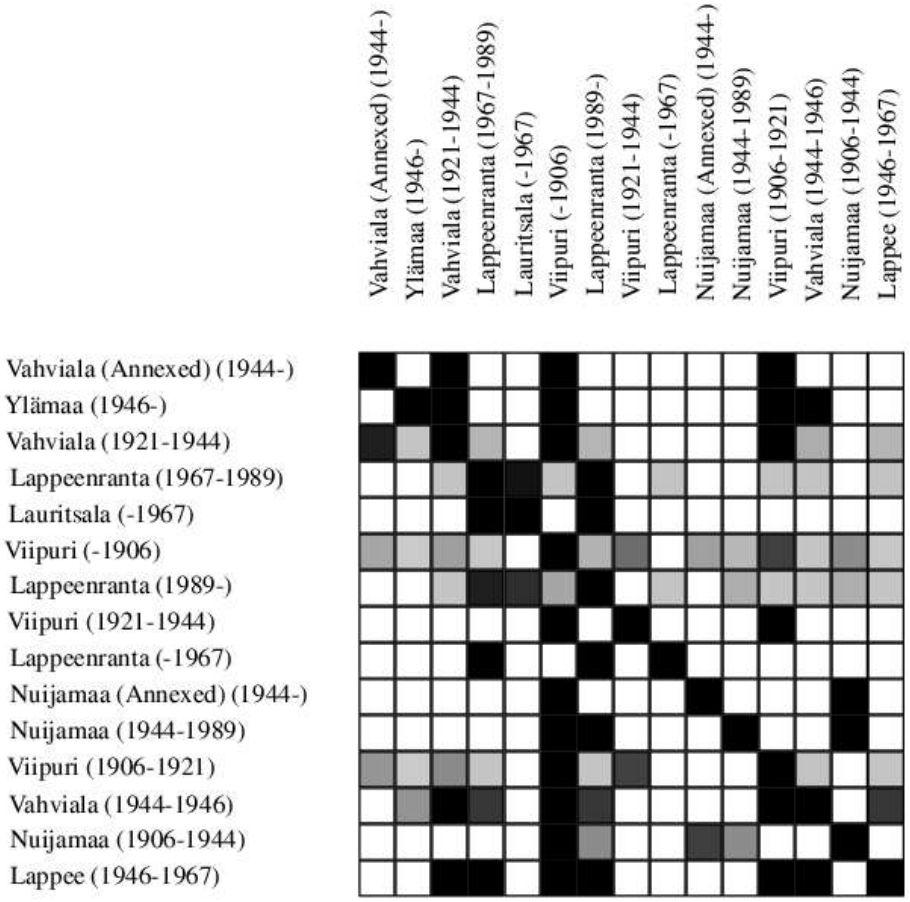
- Changes around Lappeenranta and Viipuri region from 1906 until today

An example:

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

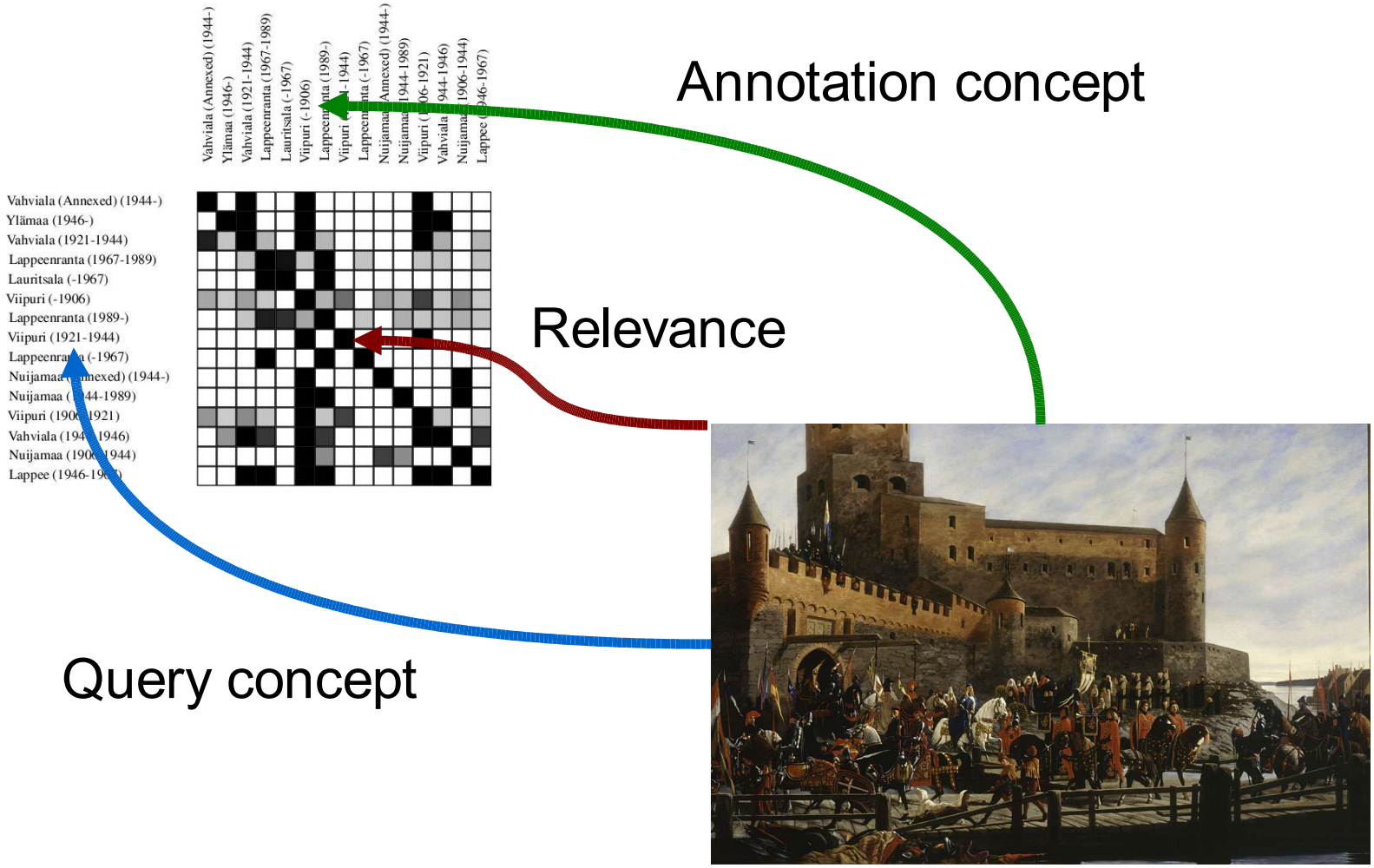


Coverages visualized

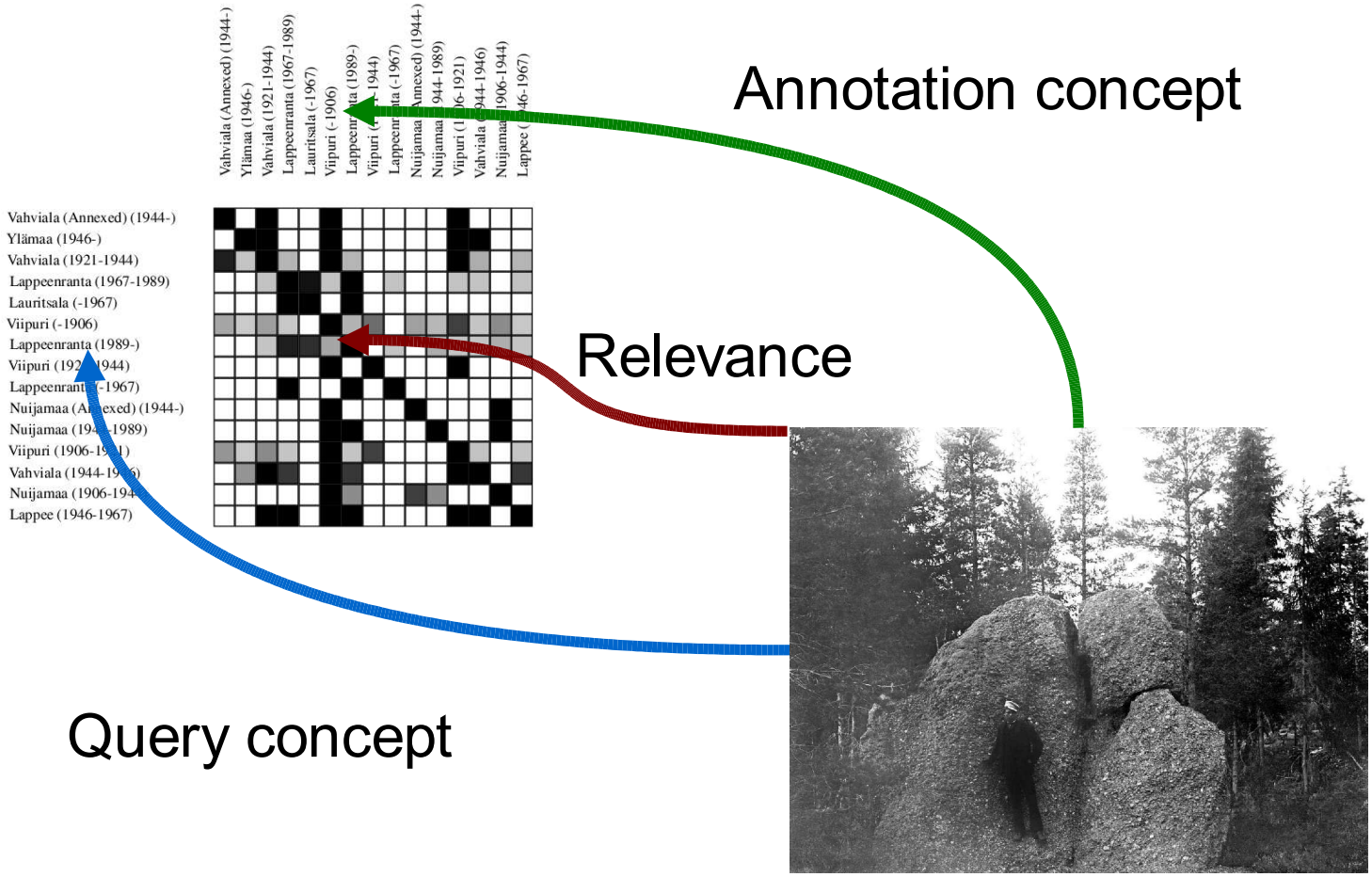


- Shades of grey indicate the **level of coverage**: the darker the box, the higher is the coverage.
- The black color indicates a full **100% coverage** between the SAPO regions and the white color a **0% coverage**.
- From this illustration it is easy to see the **mutual asymmetric coverages** between the regions

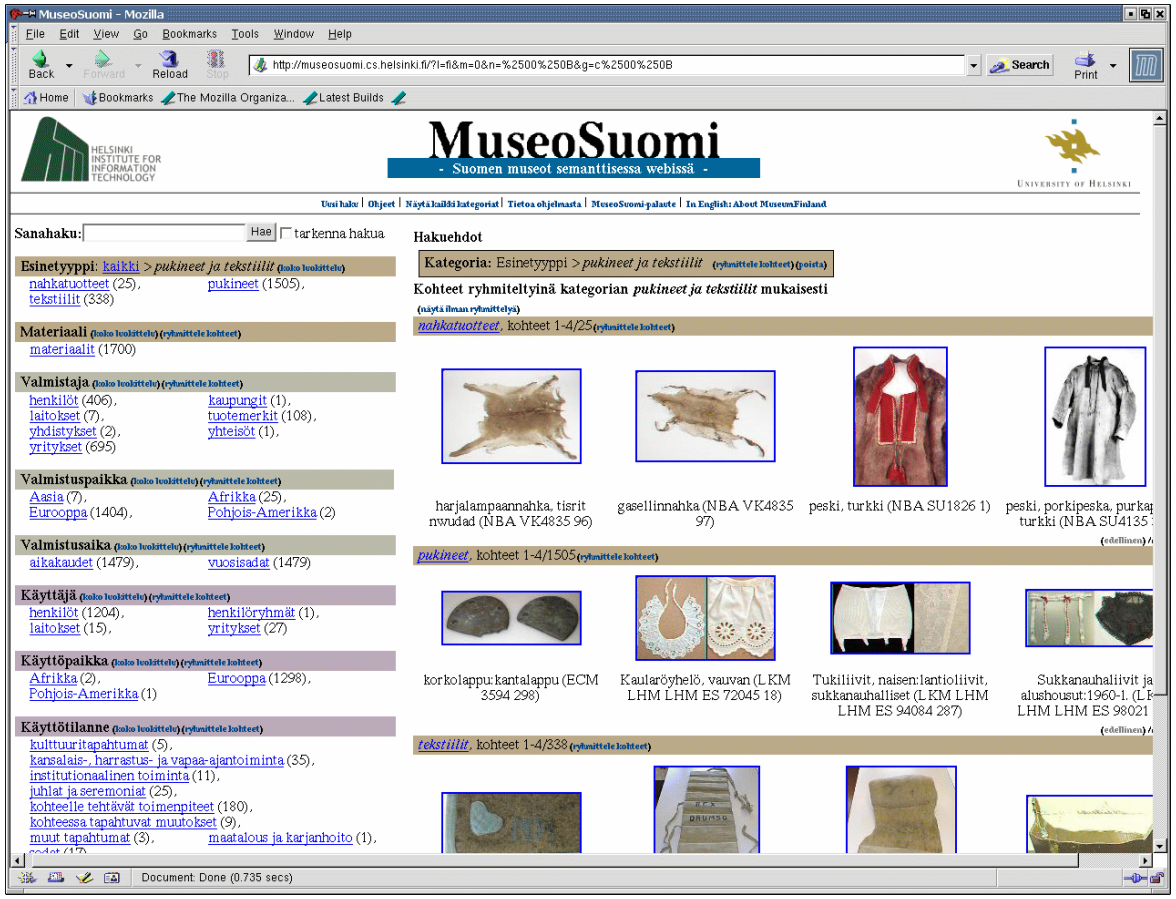
Inference results can be used in queries



Inference results can be used in queries



Applications -- MuseumFinland



Awards

Conclusions

- Ontologies are “machine-understandable” unlike words
- Better semantic content would make it possible to create **more intelligent user interfaces**
  - Typing in keywords and reading the hit list is not the only possibility!
  - Ontologies enable semantic browsing, view-based search, graphical interfaces, content visualization etc.
- Computer helps in **choosing the right concept** for content description: fancier browsing of concepts and automatic limitation of suggested concepts

Conclusions

- Ontologies provide a basis for content descriptions that is more flexible than thesauri
- Formal and exact semantics of ontologies enable the creation of intelligent applications
- Ontologies are supported by new WWW standards (Semantic Web)
    - Content publication, interoperability

Sources

- Gruber, T. R. *Toward Principles,for the Design of Ontologies Usedfor Knowledge Sharing.* Int. J. Human-Computer Studies, 43 (1995) 907-928.
- Guarino, *N. Formal Ontology and Information Systems.* Formal Ontology in Information Systems. 10s Press, 1998.
- E. Hyvönen, M. Junnila, S. Kettula, E. Mäkelä, S. Saarela, M. Salminen, A. Syreeni, A. Valo, K. Viljanen: *MuseumFinland - Finnish Museums on the Semantic Web. User's Perspective*. Proceedings of Museums and the Web 2004 (MW2004), March 31 - April 3, 2004, Arlington, Virginia, USA.

Sources

- Tomi Kauppinen and Eero Hyvönen. Geospatial Reasoning Over Ontology Changes in Time. To appear in Workshop on Spatial and Temporal Reasoning, International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, August, 2005.
- Tomi Kauppinen and Eero Hyvönen. Modeling Coverage Between Geospatial Resources. To appear in 2nd European Semantic Web Conference ESWC2005, Heraklion, Crete, May 29 - June 1, 2005.
- Mirva Salminen. Presentation slides, Helsinki, 2005.
- Eero Hyvönen. Presentation slides, Helsinki, 2005.