Finnish National Ontologies for the Semantic Web - Towards a Content and Service Infrastructure

Eero Hyvönen, Arttu Valo, Ville Komulainen, Katri Seppälä, Tomi Kauppinen, Tuukka Ruotsalo, Mirva Salminen, and Anu Ylisalmi

Helsinki University of Technology
and University of Helsinki
Semantic Computing Research Group (SeCo)
http://www.seco.tkk.fi

Content

- Why a semantic web infrastructure is needed?
  - Problem
  - Theses for the solution
- FinnONTO – Finnish Semantic Web Infrastructure Project
  - Goals
  - Solutions
  - First results

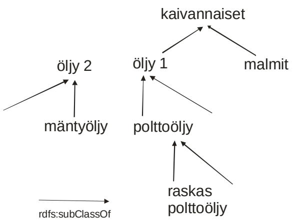
The Problem

- Four Facts
  - Semantic Web = next generation WWW
  - Ontologies = the ”silver bullet” of the Semantic Web
  - Finnish ontologies = there were none
  - Something should be done about it!
- If shared ontologies are not available, then
  - machine semantics on the web cannot be created,
  - the web will not be interoperable,
  - content creation work will be duplicated, and
  - web applications are more difficult and expensive to create.

Theses Underlying FinnONTO (1)

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- The Semantic Web needs a content infrastructure
  - like traffic needs roads
  - like energy service needs power stations & network
  - like telecommunication needs GSM etc. systems
- Major components of the infrastructure
  - shared **reference ontologies**
  - **ontology services**
  - international **standards**



Theses Underlying FinnONTO (2)

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Thesauri -> ontologies
  - Human readability -> machine understandability
- National ”ontologization” should be started
  - Content interoperability between different domains needed
  - Core ontologies should be maintained publicly
    - » Supports wide commitment to their usage
  - Core ontologies should be free and Open Source
    - » Supports wide usage and interoperability
- Business opportunities can be based on the infrastructure
  - Application specific ontologies
  - Semantic web applications

UNIVERSITY OF HELSINKI SeCo

FinnONTO-project

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Goal
  - Demonstrate benefits of a national semantic web infrastructure
  - Start an inter-organizational process towards it
- Time table
  - 9/2003-10/2005 and 10/2005-5/2007
- Organization
  - 28 public organizations & companies
    - » provide content and funding
  - 5 university labs, most research done by SeCo
- Volume
  - next year 0,8Me/year

UNIVERSITY OF HELSINKI SeCo

Three Major Goals

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- **1.** Develop open source **core ontologies**
  - General Finnish Thesaurus YSA
    -> General Finnish Ontology YSO
  - Other related ontologies
- **2.** Provide public **ontology services**
  - ONKI Server
- **3. Application demonstrations**: Semantic portals
  - Eating our own ”dog food”

UNIVERSITY OF HELSINKI
SeCo

1. Develop Core Open Source Ontologies

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Motto: Thesauri -> Ontologies
- YSO Finnish Upper Ontology
- Other related core ontologies
  - Places
    - » Including historical place name/are development
  - Actors
    - » People, companies, organizations, …
  - Domain specific ontologies
    - » Cultural concepts, photography, history, health, …

UNIVERSITY OF HELSINKI
SeCo

Case: YSO Finnish Upper Ontology

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Basis: ~23.000 terms of YSA Thesaurus, 179 domains
  - Typical thesaurus (LT/NT, RT, …)
  - Taxomies only partialle developed
  - Widely used by human indexers in different domains
- Our experience of using thesauri on the semantic web
  - Not enough: lightweight ontologies and manual work is needed!
    - » Examples will be given
  - For semantic searching and browsing

UNIVERSITY OF HELSINKI

SeCo

9

Making YSA into an ontology YSO

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- The process
  - 1. Transformation into RDF(S)/OWL
    - » Mechanical transformation into a Protege-2000 project
    - » Result: objects with following semantic properties:
      - Labels in different languages
      - Larger/Narrower term LT/NT
      - Related term RT
  - 2. Hyponymy construction by hand: solving basic semantical problems
    - » Terminology construction
    - » Disambiguating individuals from classes
    - » Disambiguating concept meanings
    - » Disambiguating LT/NT meanings
    - » Ensuring subclass transitivity
  - 3. Enriching the ontology
    - » Disambiguating RT meanings
    - » Ontology population

UNIVERSITY OF HELSINKI

SeCo

10

Terminology Construction

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Two basic choices
  - Terms attached to ontology concepts
    - » Problem:
      - lots of terms may refer to a concept
        - E.g. different languages
      - term usage may be context dependent
        - E.g. conventions used in organizations
  - Separate modular terminologies (our choice)
    - » Problem:
      - managing different terminologies

UNIVERSITY OF HELSINKI SeCo

11

Disambiguating Individuals from Classes

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- An example from YSA thesaurus
  - ”Halley comet” LT ”comet”
  - Problem: ”Halley comet” is an individual of the class ”comet” but the machine does not know this
    - » There can be many comets with properties inherited from the class ”comet” (and its super classes)
    - » There cannot be several ”Halley comets”
    - » An individual cannot belong to several classes but classes can
      - This is a convention e.g. in Protege

UNIVERSITY OF HELSINKI SeCo

12

Disambiguating Concepts

- An example from YSA thesaurus
  - Term ”johtaminen” (directing) can mean
    - » Directing an organization
    - » Directing a musical performance
    - » Leading electricity
    - » Leading to some place (e.g. a road)
    - » …
  - Problem: semantic confusion for the machine
    - » E.g., what to retrieve for keyword ”johtaminen”
  - Solution: represent different meanings by different concepts
    - » At least major semantic distinctions should be made

Disambiguating LT/NT Meanings

- LT/NT relation has many meanings
  - the machine cannot disambiguate them although humans can
- An example from YSA thesaurus
  - ”hospital” LT ”health care institution”
    ”comet” LT ”solar system”
  - Problem of semantical confusion:
    - » ”hospital” subClassOf ”health care institution”
      - OK: hospitals inherit properties of health care institutions
    - » ”comet” subClassOf ”solar system”
      - WRONG: a comet is not a solar system and does not have own planets etc.
- Solution:
  - ”hospital” subClassOf ”health care institution”
    ”comet” partOf ”solar system”

Ensuring Subclass Transitivity

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- An example from YSA thesaurus
  - ”fuel oil” LT ”oil”
    ”pine oil” LT ”oil”
    ”oil” LT ”diggings”
  - Problem: ”pine oil” is not ”diggings” but ”fuel oil” is
    - » The machine cannot understand the difference although humans can
- Solution
  - Differentiate multiple meanings of ”oil”
  - Create two subClassOf hierarchies

UNIVERSITY OF HELSINKI SeCo 15

Disambiguating RT relations

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Example from YSA:
  - ”Solar wind” RT ”Northern lights”
- Solution approach:
  - Identify a small set of generic role relations
  - Use frames (cf. e.g. FrameNet):
    - Events are of central importance
    - Event types can be characterized by a set of roles
      - E.g. selling/buying event may have roles Agent, Recipient, Location
      - Terms sell/buy refer to the same event
- For example:
  - ”Solar wind” RT ”Northern lights”
    -> ”causes1” source ”Solar wind”
    result ”Northern lights”
- This work will be first tested in domain specific applications

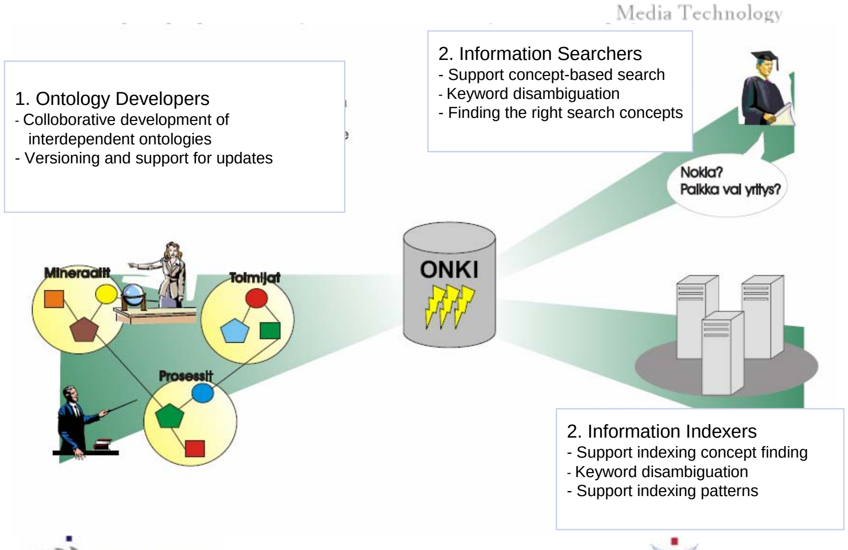
UNIVERSITY OF HELSINKI SeCo 16

2. Ontology Server ONKI

- Problem
  - A single authority cannot manage large multi-domain ontologies
    - » E.g., maintaining YSA is difficult
  - Ontologies have to be developed collaboratively by domain expert groups
    - » However, this leads to organizational difficulties
- Support is needed for
  - distributed **development**,
  - **versioning**,
  - **publishing**, and
  - **usage** through the web.
- Solution: ONKI server with its services
  - Support development and publishing processes
  - Support usage as web services

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

ONKI Services & User Groups



Collaborative Ontology Development in ONKI

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Basic workflow
  - Each inter-related ontology in maintained by a domain expert group DEG
    - » DEG maintains its own concepts as an ontology
    - » DEG publishes new versions of it in ONKI
    - » Each concept is ”owned” by one DEG
  - Other groups can ”borrow” concepts from their owners by using ”proxies”
    - » Properties can be inherited from the home ontology
    - » Concepts can be modified locally
  - When an owner of a concept modifies her ontology
    - » changes are documented as instances of changes (an ontology)
    - » changes are stored in ONKI when a new version is committed there
  - When a borrower wants to update his ontology to match a new version of a dependent ontology
    - » change history can point out possible problems to her and
    - » can help her in making the needed updates
- Design and implementation is underway

UNIVERSITY OF HELSINKI SeCo

19

ONKI Browser: A Use Example

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

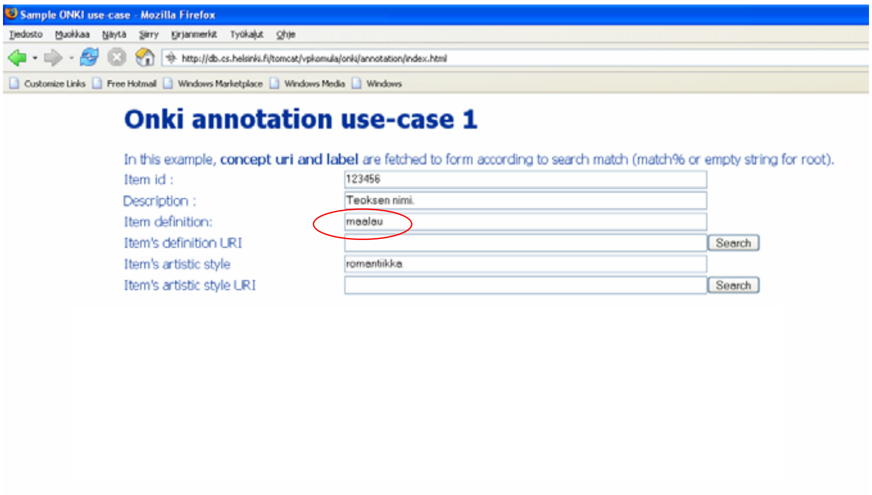
- **Setting**
  - **A legacy cataloging system in a museum**
  - **The museum is convinced that cultural ontologies are needed for indexing collection artifacts**
  - **”How can be start producing semantic metadata in practice?”**
- **Problem:**
  - **How to *find* right up-to-date labels, URIs, etc. for metadata?**
  - **How to *transfer* them easily into the database?**
- **Solution:**
  - **Using ONKI as a web service**
    - **» Search & browsing functions**
    - **» Copying URIs etc. data into the local system**

UNIVERSITY OF HELSINKI SeCo

20

Step 1: Typing in ”maalau..”

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology



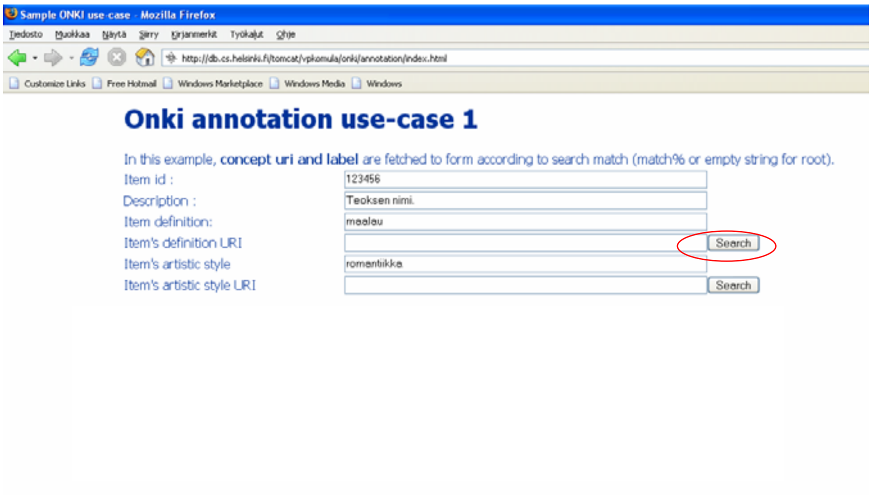
UNIVERSITY OF HELSINKI

SeCo

21

Step 2: Autocompleting the input semantically (search)

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology



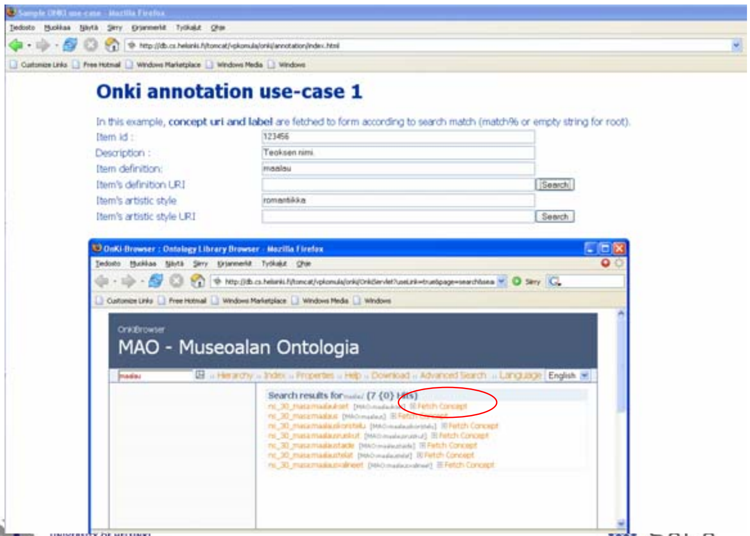
UNIVERSITY OF HELSINKI

SeCo

22

Step 3: ONKI Browser opens with matching MAO concepts: select one

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology



UNIVERSITY OF HELSINKI

SeCo

23

Step 4: Label & URI copied into the cataloging system

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

UNIVERSITY OF HELSINKI

SeCo

24

3. Pilot Applications

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- eCulture
  - MuseumFinland
    – Finnish Museums on the Semantic Web
    - » http://www.museosuomi.fi
  - CultureSampo
    – Finnish Culture on the Semantic Web
- eGovernment
  - Suomi.fi
    - » http://www.museosuomi.fi/suomifi
- eHealth
  - National health portal
  - Sosiaaliportti.fi
- eLearning
  - YLE Kaffi (2000 video clips + 4500 artifacts)
    - » http://www.museosuomi.fi/orava
  - Opintoluotsi.fi (educational resources)
  - Tiedonportti.fi (library resources)

A new application award
Semantic Web Challenge
http://challenge.semanticweb.org

nordic digital excellence in museums
nodem forum & award

Online semantic portal demos

UNIVERSITY OF HELSINKI

SeCo

25

Thank you

HELSINKI UNIVERSITY OF TECHNOLOGY
Media Technology

- Questions?

UNIVERSITY OF HELSINKI

SeCo

26