

Semanttinen web kansalliseksi voimavaraksi
FinnONTO(2003-2007) hankkeen visio ja tulokset
25th January 2008

QUCCOO & ShOE
Tools for building search ontologies and
ontology-based searching in unannotated
text collections

Feza BAŞKAYA

University of Tampere
Department of Information Studies

Outline

- 1. Background
- 2. Ontologies
- 3. Quccoo: Searching Unannotated Text Collections through Ontologies
- 4. ShOE: Creating ontologies
- 5. Discussion, Conclusion

1. Background

- Vast online information environments
 - many different natural languages
- Vocabulary mismatch
 - hard to guess the best search keys; leads to loss of search effectiveness
 - especially in foreign languages
 - hard to know word forms, compound treatment
- Other problems – depending on one's search environment
 - collection dependency, metadata dependency
 - engine and query language dependency

2. Ontologies

- Ontologies model semantics
 - concepts
 - rich relationships
 - support inference
 - application means resource annotation
 - closely related to thesauri
- Ontologies can solve the vocabulary problem!
 - represents the semantics of resources (documents) better than pure natural language
 - retrieval becomes correct and accurate
 - desired: a universal world model, and a controlled language for description and reasoning about this model

Issues in Classification and Indexing

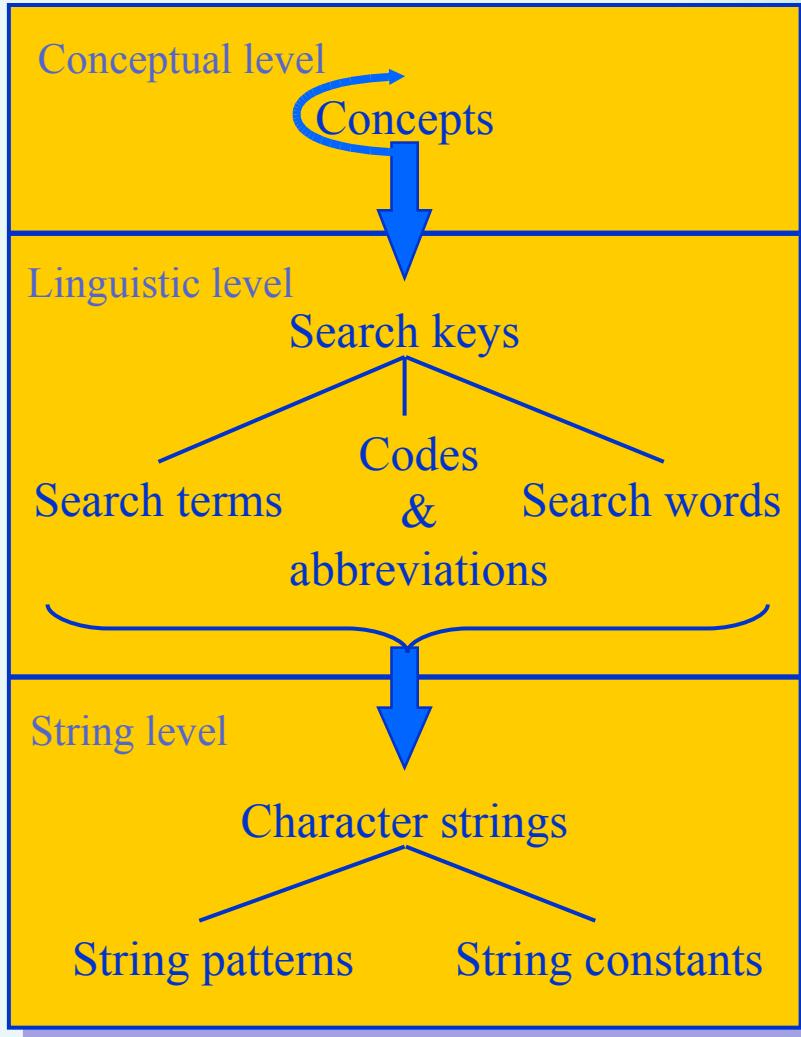
- Index languages -
 - modeling - coverage, viewpoint
 - maintenance - ageing, cost
- Indexing -
 - specificity, exhaustivity, consistency
 - cost - where paid, who pays?
 - The over-specificity the devices created often lead to poor recall and thus they were soon mostly abandoned

Any Room for Ontologies?

3. Searching Unannotated Text Collections through Ontologies

- Searching ontologies can
 - provide conceptual organization
 - support direct access to textual content
 - ✓ translate between concepts and textual variation
 - ✓ translate between natural languages
 - ✓ hide search engines / query languages
 - ✓ may support *other* media / structures / features
 - be light-weight, narrow, and no world models
 - ✓ personal, group or small community support
 - ✓ versions, mutually incoherent, easily modifiable
 - ✓ easily disposable, perhaps tradable

Three levels



➤ Forest industry

- forest industry
- paper industry
- saw mill
- ...

- pl(saw, mill)
- al(industry)
- pl(paperi, tehdas)

QUCCOO: Principles

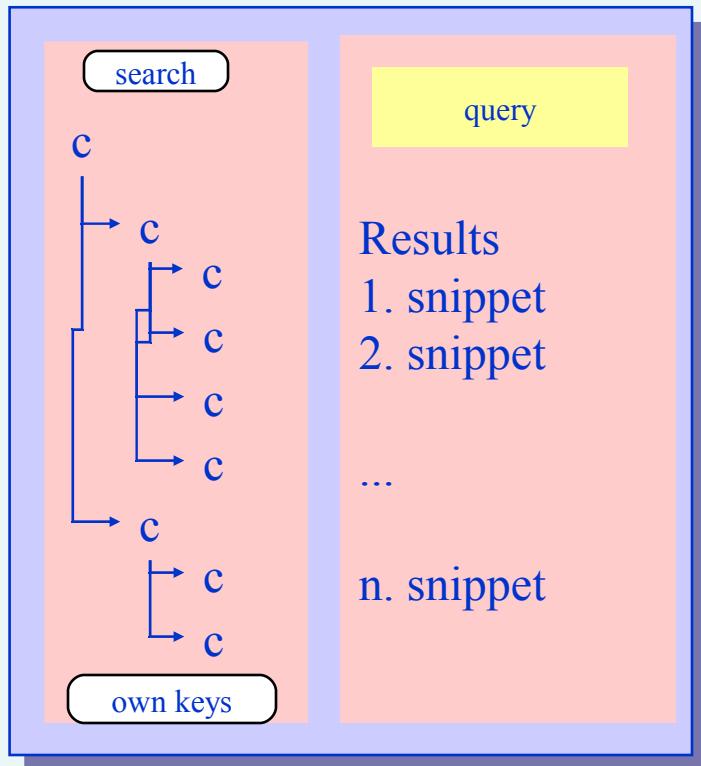
- QUCCOO: QUery ConstruCtion with OntOlogies for direct content access
- Based on the three levels ...
- Aims to provide *independence* of ...
 - expression variability (nutraceutical?)
 - natural language (French?)
 - collection (intranet, Web ,....)
 - indexing (lemmatization, compounds?)
 - availability of metadata & world model
 - engine & query language (Lemur, Trip, Google, ...)
- You just select your concepts, targets and go!
 - Point, click and go

QUCCOO: Status

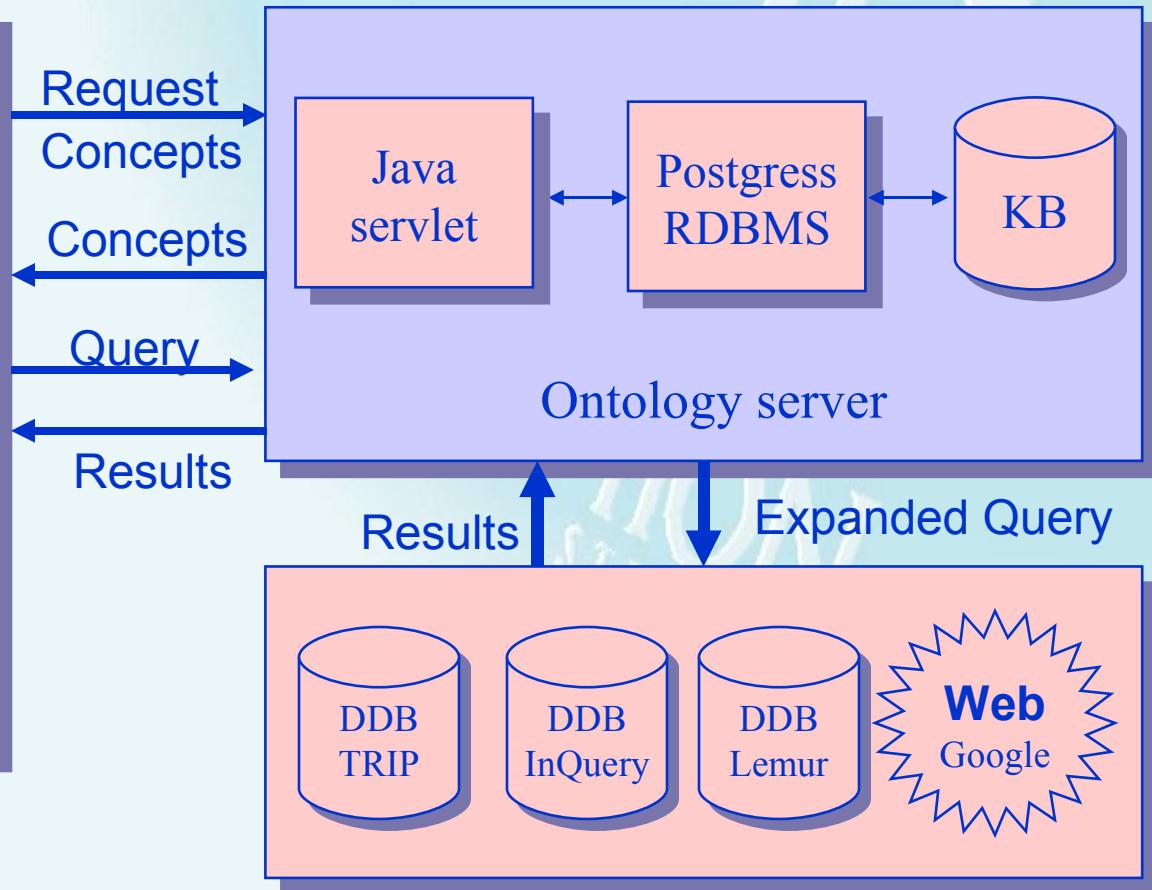
- Web application, uses state-of-the-art Servlet technology
- Supports diverse full-text database engines (Trip, InQuery, etc.) as well web search engines (e.g., Google)
- Supports diverse collections
- Intuitive; simple interface to access information
- Supports multilingual search and various index types

QUCCOO: Architecture

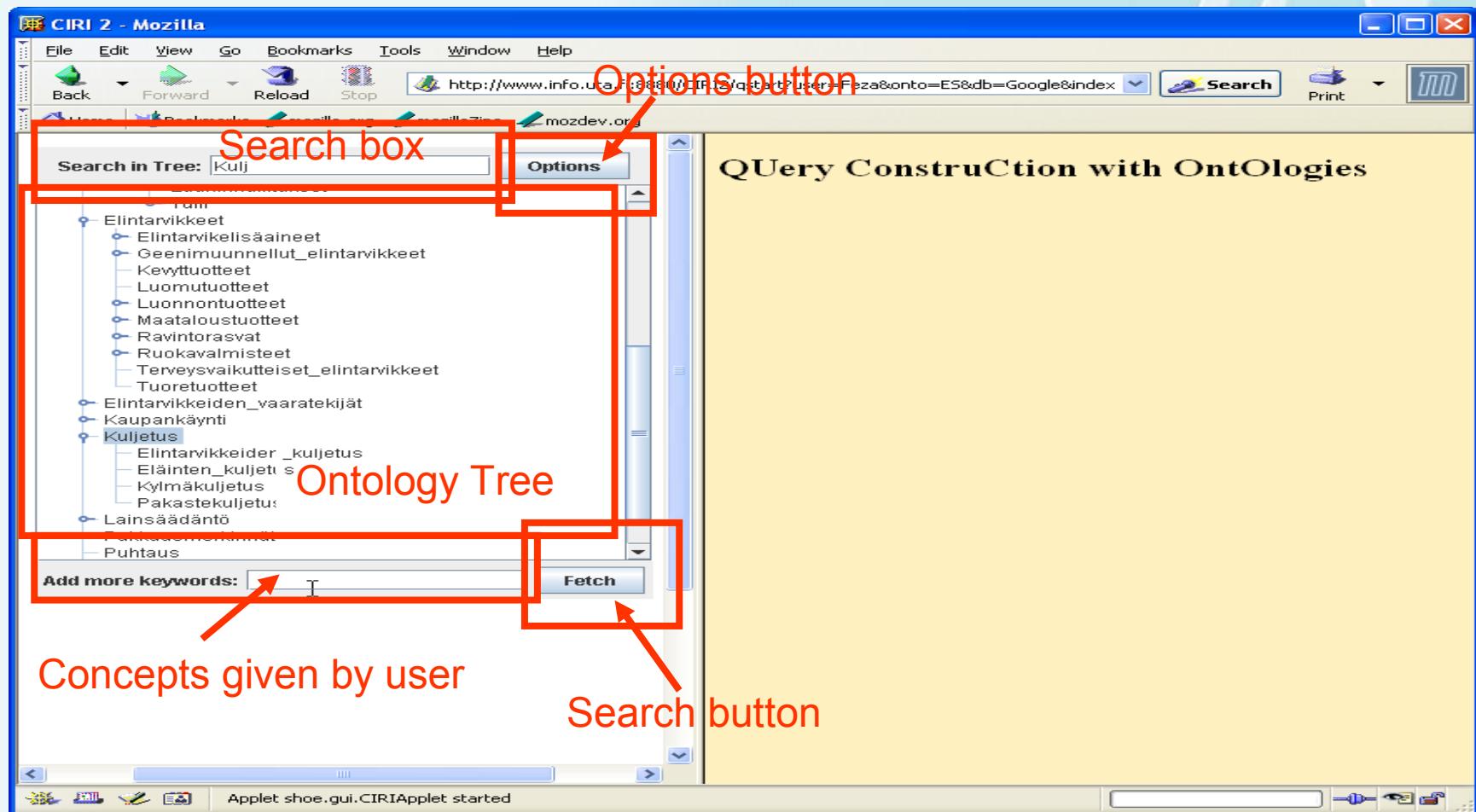
Client- Applet



Server-Servlet



Quccoo - interface



Quccoo - interface

Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.info.uta.fi:8880/WebOntoClassification/ClikObserver.html

Mail :: Inbox (...) Sun Statement (J...) Getting Data ... LEO D-E Erge... http://...e.html phpPgAdmin Index of /feza http://...html

Search in Ontology: Hide ...
Add more keywords: Fetch
Search Engine: Lemur Database: STT:bas...
Liberality: Liberal Query Lang.: fi
Expansion Le...: 1 User name:
Ontologies: Ontologies Onto Lang.: Finnish

Ontology

- Elintarvikealan_käsitteet
- Ympäristöontologia
 - Jätteet
 - Luonnonvarat
 - Luonnonvarojen_käyttö
 - Uusiuutumattomat_luonnonvarat
 - Uusiuutuvat_luonnonvarat
 - Ympäristön_tila
 - Ympäristöindikaattorit
 - Ympäristön_seuranta
 - Ympäristötilastot
 - Tutkimushankkeet
 - Tutkimushankkeet
 - Tutkimusverkostot
 - Ympäristönsuojelu
 - Ilmansuojelu
 - Ilmastonsuojelu
 - Luonnonsuojelu
 - Harjujen_suojelu
 - Laijensuojelu
 - Luontotyppien_suojelu
 - Metsiensuojelu
 - Soiden_suojelu
 - Vesiensuojelu



Lemur Search

0: Eduskunta (Twolloed)

#sum(#syn(ympäristönsuojelu #uw12(ympäristö suoje...))

Search

Searched for #sum(#syn(ympäristönsuojelu #uw12(ympäristö suoje...)) ilmans

Käyttäjätunnus:

1. [SLL: Hallitus ei ole toistaiseksi täyttänyt ympäristölupaauksiaan](#)

(Score: 0.625408)

2. [SLL: Hallitus ei ole toistaiseksi täyttänyt ympäristölupaauksiaan](#)

(Score: 0.625209)

3. [Luonnonsuojeliitto vaati lisää soita suojeleun](#)

(Score: 0.624323)

Quccoo - interface

The screenshot shows a Mozilla Firefox browser window titled "CIRI 2 - Mozilla". The address bar contains the URL <http://www.info.uta.fi:8880/CIRI2/qstart?user=Feza&onto=ES&db=Google&index=basic&dlang=finnish&qlang=en&se>. The main content area displays search results from the "AAMULEHTI ARKISTO" database. The results are titled "HAKUTULOS" and show 2376 documents found. The first few results are listed:

1. Dollarikin nostaa raakaöljyn hintaa Ei arviota
2. Palvelualojen liikevaihto syksyllä neljän prosentin kasvussa Ei arviota
3. Yrittäjät odottavat kunnalta strategiaa palveluseteistä Ei arviota
4. Posti hakee uutta bisnestä sähköomun kierrätyksestä Ei arviota
5. Taksien kysyntä hiipuu yhä Ei arviota
6. VR Osakeyhtiö Ei arviota
7. Transbox Oy Ei arviota
8. Kelpo Kuljetus Fi Oy Ei arviota
9. EU:n komissio pani rajat halpalentoihioiden tulle

To the left of the search results, there is a sidebar titled "Search in Tree" containing a hierarchical tree view of an ontology. The tree includes categories like "Lamminmäkimerkset", "Tulli", "Elintarvikkeet", "Elintarvikkeiden_vaaratekijät", "Kaupankäynnit", "Kuljetus", and "Lainsääädäntö". A "Fetch" button is located at the bottom of this sidebar.

Ontology in Finnish

Trip Database Engine results

Quccoo - interface

The screenshot shows a Mozilla Firefox browser window titled "CIRI 2 - Mozilla". On the left, there is a sidebar titled "Search in Tree" containing a hierarchical tree structure of terms related to "Kuljetus" (Transportation). The tree includes categories like "Laajimmamukiset", "Tulli", "Elintarvikkeet", "Geenimuunnellut_elintarvikkeet", "Kevytluotteet", "Luomuluotteet", "Maataloustuotteet", "Ravintorasyt", "Ruokavalimisteet", "Terveysvalkuutteiset_elintarvikkeet", "Tuoretuotteet", "Elintarvikkeiden_vaaratekijät", "Kaupankäynti", "Kuljetus", "Elintarvikkeiden_kuljetus", "Eläinten_kuljetus", "Kylmäkuljetus", "Pakastekuljetus", and "Lainsääädäntö". Below the tree is a search bar with "Add more keywords:" and a "Fetch" button. The main content area displays Google search results for the query "((kuljetus))". The results are in Finnish and include links to various companies like Teljän Kuljetus Oy, Tonttilan Nosto ja Kuljetus Oy - Etusivu, Kuljetus H. Valo Oy, Skal - WWW Etusivu, www.savonkuljetus.fi/, PH Kuljetus Oy, Kuljetus - huolinta - Venäjä - transito - kuljetusliikkeet ..., Kelpo Kuljetus Fi Oy, and others. The results are dated Nov 2005. A sidebar on the right shows "Sponsored Links" for LKW Walter - kuljetus and Need Kuljetus.

Ontology in Finnish

Google Results in Finnish

Quccoo - interface

CIRI 2 - Mozilla

File Edit View Go Bookmarks Tools Window Help

Back Forward Reload Stop http://www.info.uta.fi:8880/CIRI2/qstart?user=Feza&onto=ES&db=Google&index=basic&dlang=finnish&qlang=eng

Search Print

Home Bookmarks mozilla.org mozillaZine mozdev.org

Search in Tree: Kulj Hide ...

Search Engine: Google Query Language: en

Kuljetus

- Elintarvikkeet
 - Tulli
 - Elintarvikkeet
 - Elintarvikeliisääneet
 - Geenimuunnellut_elintarvikkeet
 - Kevyttooteet
 - Luomutooteet
 - Luonnontuoteet
 - Maataloustuoteet
 - Ravintorasyt
 - Ruokavalimisteet
 - Terveysvaikuttavat_elintarvikkeet
 - Tuoretuoteet
 - Kaupankäynti
 - Kuljetus
 - Elintarvikkeiden_kuljetus
 - Eläinten_kuljetus
 - Kylmäkuljetus
 - Pakastekuljetus
 - Lainsäädäntö

Add more keywords: Fetch

Google Web Images Groups News Froogle Local more » Advanced Search Preferences

Results 1 - 10 of about 690,000,000 for ((transport OR transit OR transportation OR convey OR carrying OR conveying OR haulage OR hauling))

News results for ((transport OR transit OR transportation OR conveyance OR carrying OR conveying OR haulage OR hauling)) - View today's top stories

Codey Is Seeking to Replenish State's Transportation Fund - New York Times - 8 hours ago

Bomb threats freeze Brisbane transport - NEWS.com.au - 16 hours ago

High-tech, high transport security? - BBC News - 18 hours ago

Transport for London homepage - [Transport for London](#)

Information on all forms of public **transport** in London, routes, maps, journey planner and online tickets sales. Includes guide to what to see and do and ...

[www.tfl.gov.uk/](#) - 19k - 13 Nov 2005 - [Cached](#) - [Similar pages](#)

US Department of Transportation

Oversees federal highway, air, railroad, and maritime and other **transportation** administration functions; components include the FAA, FHA, FRA, NHTSA, OIG, ...

[www.dot.gov/](#) - 24k - 13 Nov 2005 - [Cached](#) - [Similar pages](#)

New Jersey Transit

NJ **TRANSIT** is New Jersey's public **transportation** corporation. Its mission is to provide safe, reliable, convenient and cost-effective **transit** service with a ...

[www.njtransit.com/](#) - 34k - 13 Nov 2005 - [Cached](#) - [Similar pages](#)

Bureau of Transportation Statistics (BTS) - Research and ...

Compiles, analyzes, and makes accessible information on the Nation's **transportation** systems.

[www.bts.gov/](#) - 29k - 13 Nov 2005 - [Cached](#) - [Similar pages](#)

National Transportation Safety Board

The National **Transportation** Safety Board is an independent Federal agency charged by Congress with investigating every civil aviation accident in the United ...

[www.ntsb.gov/](#) - 17k - 13 Nov 2005 - [Cached](#) - [Similar pages](#)

Transportation Research Board -- Homepage

Transportation Cost Implications of New Development Proposals Due November 15, 2005 ... Guide for **Transportation** Landscape and Environmental Design ...

[www.trb.org/](#) - 32k - 13 Nov 2005 - [Cached](#) - [Similar pages](#)

Ontology in Finnish

Google Results in English

Quccoo - interface

The screenshot shows a Mozilla Firefox browser window with the title "CIRI 2 - Mozilla". On the left side, there is a sidebar titled "Search in Tree" containing a tree view of keywords. The tree includes categories like "Laamimmuukset", "Tulli", "Elintarvikkeet", "Geenimuunnellut_elintarvikkeet", "Kevyttootteet", "Luomutootteet", "Lounantuooteet", "Maataloustuotteet", "Ravintorasyat", "Ruokavalimisteet", "Terveysevaikuttajiset_elintarvikkeet", "Tuoretuotteet", "Elintarvikkeiden_vaaratekijät", "Kaupankäynti", "Kuljetus", "Elintarvikkeiden_kuljetus", "Elainten_kuljetus", "Kylmäkuljetus", "Pakastekuljetus", and "Lainsääädäntö". Below the tree is a text input field "Add more keywords:" and a "Fetch" button. On the right side, the main content area shows a Google search results page for the query "((transport OR forslig OR fraktnings OR befordran OR överföring OR befordran))". The results are displayed in Swedish, with titles like "Electronic Transport in Mesoscopic Systems", "Fundamentals of Carrier Transport", "Transport and Developing Countries", "Transport for London homepage", "London Underground homepage", "Transport Canada - Transports Canada", "Department for Transport", "International Air Transport Association", and "US Department of Transportation". Each result includes a snippet of text and links to the cached and similar pages.

Ontology in Finnish

Google Results in Swedish

Quccoo - interface

The screenshot shows a Mozilla Firefox browser window titled "CIRI 2 - Mozilla". The address bar displays the URL: <http://www.info.uta.fi:8880/CIRI2/qstart?user=Feza&onto=ES&db=Google&index=basic&dlang=finnish&qlang=en>. The left panel contains a search interface with a tree view of search terms under "Search in Tree: Kulj". The tree includes categories like "Elintarvikkeet", "Kuljetus", and "Lainsääädäntö". A "Fetch" button is at the bottom of this panel. The right panel shows Google search results for "Kuljetus" with approximately 70,700 results. The results list various companies like "Teljän Kuljetus Oy", "Kuljetus H. Valo Oy", and "Kuljetus - huolinta - Venäjä - transito - kuljetusliikkeet ...". The entire interface has a blue and white color scheme.

Ontology in Finnish

Google Results with Logging Facility

4. ShOE: Creating ontologies

- **Search Ontology Editor** - for creating ontologies
 - supports the 3 layer architecture of QUCCOO
 - intuitive; easy to learn and use
 - automatic support for the human editor
- Multilingual in many aspects
 - GUI, User Interface language can be changed
 - Concepts names can be edited/displayed on-the-fly in different languages
 - Expressions can be edited/displayed on-the-fly in different languages.

ShOE: Implementation

- Well-designed modular object-oriented architecture based on MVC paradigm
- Platform independent; written in Java
- Flexible; e.g. uses XML as file format, configurable tables, with XML configurable menu structure
- Robust
- Extensible via Plug-ins

ShOE - Main window

Search field

Concept hierarchy tree

Concept properties

Description

Expressions Assosiations

Expressions window

| Name | Etype | Strength | Occurrences |
|-----------------------|-------|----------|----------------------------|
| yritys | term | 1 | Click here |
| elintarvikealan_yr... | syno | 0,25 | Click here |

Add Expression Delete Expression

6. Conclusion

- ShOE and QUCCOO are *one* answer to problems in semantic information access
 - light-weight disposable search ontologies for full content access
 - independencies of:
 - ✓ collections (partially), indexing ways,
 - ✓ availability of metadata / annotations
 - ✓ changes of needs, variability of "world models"
 - ✓ search engines, query languages
 - ✓ vocabulary variation and natural languages
 - a compromise, different from semantic annotation or indexing, with control at the user end

User testing

Cross-language Web search

- Test persons
 - 40 students from the University of Tampere and Pirkanmaa polytechnic
- Ontology
 - Combination of two ontologies: Food concepts and geographical concepts
- 2 interfaces
 - QUCCOO + interface without ontology (basic Google search)
- 4 simulated search tasks
 - Two tasks with one interface and two with the other

Analysis

- Log files
 - Queries
 - Relevance assessments (scale 0-4)
- Questionnaires
 - Opinions about ontology and Quccoo-interface

Results: search success

- No significant difference between systems
 - QUCCOO performed better when strong query structure was needed (“alcoholic beverage”)
 - In most self-formulated queries no phrases were used
- QUCCOO helps persons who are not used to formulate structured queries

Results: opinions

- "Structure of the ontology was logical"
- "Finding search concepts needed in the tasks in ontology was easy"
- "Using the ontology was effortless"
 - 92 % agreed in all

Results: opinions

- 32/40 thought that QUCCOO-interface was easier to use
- 32/40 liked QUCCOO better
- Why?
 - Helped users to clarify task topic and to find related search keys
 - Made cross-language search easy (in 80% of direct searches some dictionary was used to help query formulation)

Semanttinen web kansalliseksi voimavaraksi
FinnONTO(2003-2007) hankkeen visio ja tulokset
25th January 2008

OntoClass
Ontology based Classification
System
A tool for automatic document
classification

Feza BAŞKAYA

University of Tampere
Department of Information Studies

Ontoclass, Ontology based Classification System

Classify text documents

Standalone and web versions

Cross-language classification capability

Visual presentation of the classification results

easy to comprehend and easy to interpret

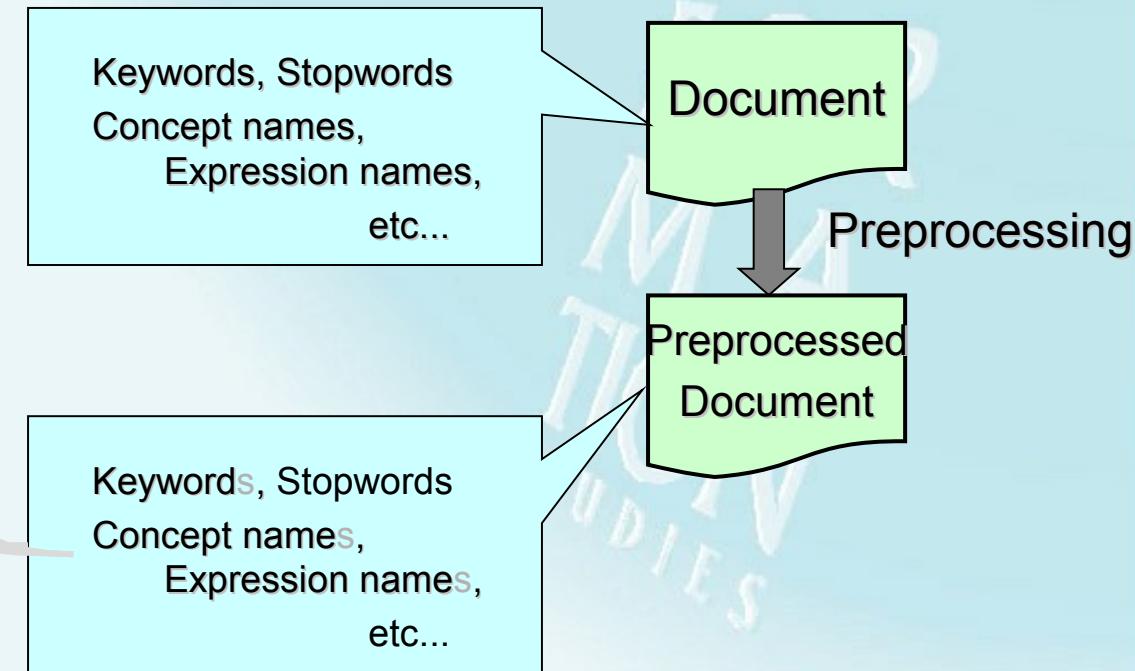
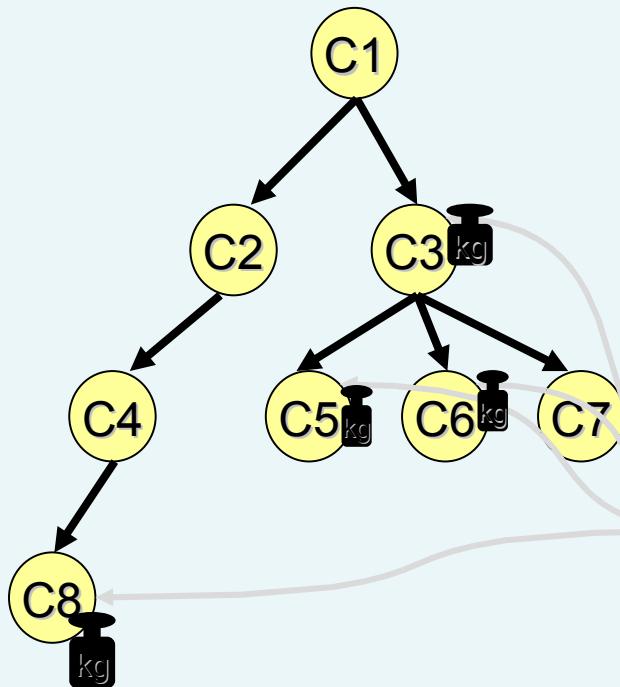
based on the visual comprehension faculty of the user

Ontoclass, Ontology based Classification System

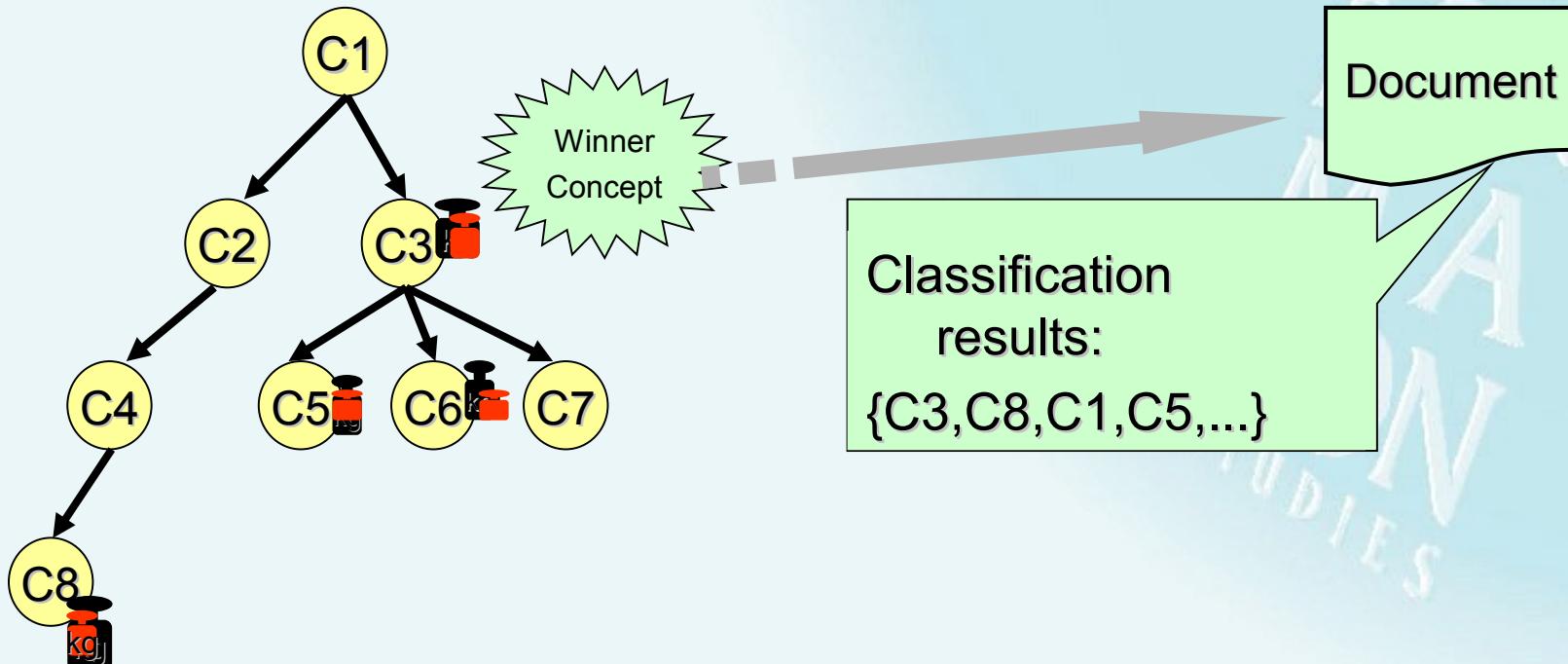
Consists of the following processes:

- Text preprocessing (lemmatizing stemming, etc.)
- Mapping the document to Ontology
- Weight propagation in ontology tree
- Selecting Winner concept(s)

Ontoclass, Ontology based Classification System



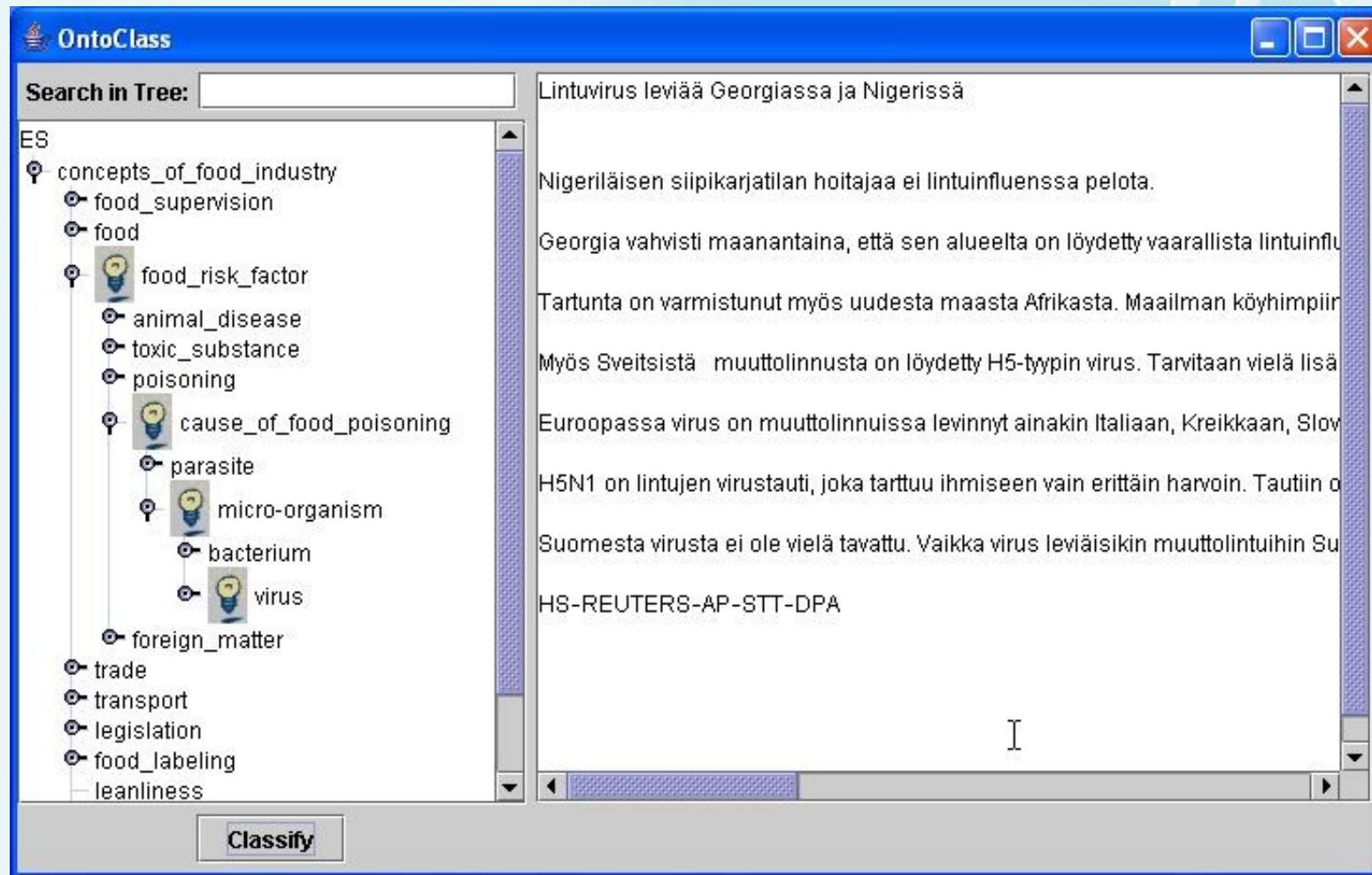
Ontoclass, Ontology based Classification System



Weight propagation

OntoClass

User interface (Standalone version)



OntoClass (Web interface)

Search in Ontology: Hide ...

Add more keywords: Fetch

Search Engine: GoogleCl... Database: STT:bas...

Liberality: Liberal Query Lang.: fi

Expansion Le...: 1 User name:

Ontologies: Ontologies Onto Lang.: English

Ontology

- concepts_of_food_industry
 - food_supervision
 - food
 - food_additive
 - gene-altered_food
 - light_product
 - organic_food
 - wholefood
 - agricultural_product
 - food_fat
 - food_product
 - functional_food
 - perishables
- food_risk_factor
 - animal_disease
 - toxic_substance
 - poisoning
 - cause_of_food_poisoning
 - foreign_matter
- trade
- transport
- legislation

Lähes luomutuote - Elintarviketurvallisusvirasto Evira

Lähes luomutuote. Jos viljelijä on siirtymässä tavanomaisesta luomutuotantakin kasviperäisissä tuotteissa, joita on jo viljely luomusäännösten mukaan ...
<http://www.evira.fi/portal/fi/elintarvikkeet/elintarviketietoja/luomuelintarvikkeet/l>

MySpaceTV Videos: Luomutuote Video Channel

Upload and share videos, find and watch funny clips, vote and comment on your favorites, subscribe to online video channels, add cool videos to your MySpace profile.
<http://vids.myspace.com/index.cfm?fuseaction=vids.channel&ChannelID=220>

luomutuote.com - luomutuote

luomutuote.com - luomutuote ... LUOMUTUOTE.COM. Tervetuloa luomutuote.com - luomutuote.com - luomutuote.com

Vinkkilän Luomutuote

Vinkkilän Luomutuote on erikoistunut terveelliseen täsmäravintoon. Tynni on Vinkkilän päätuote, mutta sen lisäksi viljelemme parsaa, mustikkaa ja aroniaa.
<http://www.vinkkilanluomutuote.fi/-15k>

Vinkkilän Luomutuote

Vinkkilän Luomutuote ... Tautien hoito kasvisravinnolla · Vanheneminen · Vitamiinit. Copyright © 2007 Vinkkilän Luomutuote. Designed by Heikki Liu
<http://www.vinkkilanluomutuote.fi/terveystieto.html> -257k -

Luomutuoteet - Luomutuote - Elintarvike - Ektuote - Aromaterapiat ...

Luomutuote on luonnonmukaisesti viljelyt. Luomutuote - Elintarvikkeet - Että Valitse Luomu tästä! Maksa verkopankissa; Nordean Solon, ...
<http://www.luomukauppa.fi/verkkokauppa/> -4k -

Iso-Markun luomutila

Joulukuusset · Perunat · Reseptit · Utiset. Etsi sivulta. Arkisto. January 2007 - December 2007. Sähköposti: [luomutuote \(ät\) luomutuote \(piste\) net](mailto:luomutuote@luomutuote.net).
<http://www.luomutuote.net/> -11k -

Luomutuote

email: info@luomutuote.fi. Ajankohtaista. Luomutuote studiossa 6.8.2007 kuuntelasmaassa täältä: www.myspace.com/luomutuoteband. Seuraava keikka: <http://www.luomutuote.fi/> -4k -

MySpace.com - Luomutuote - Espoo, FI - Rock / Punk / Alternative ...

MySpace music profile for Luomutuote with tour dates, songs, videos, pictures, fans, and more.

Search in Tree:

Ontology: <http://www.info.uta.fi:8880/QuccooFlexServer/YO.xml>

Concept language: Finnish

Ontology

- Ympäristöontologia
 - Jätteet
 - Luonnonvarat
 - Luonnonvarojen_käyttö
 - Uusiutumattomat_luonnonvarat
 - Uusiutuvat_luonnonvarat
 - Metsävarat
 - Muut_uusiutuvat_luonnonvarat
 - Kalat
 - Marjat
 - Porot
 - Riista
 - Sienet
 - Turve
 - Uusiutuva_energia
 - Vesivarat
 - Viljell...
 - Ympäristön_tila

OntoClass (Web interface)

The screenshot shows the OntoClass web interface running in Mozilla Firefox. The browser window has a title bar for 'Mozilla Firefox' and a URL bar showing 'http://www.info.uta.fi:8880/QuccooFlexServer/ClikObserver.html'. The main content area displays search results and an ontology tree.

Search in Ontology: Hide ...
Add more keywords: Fetch
Search Engine: GoogleC... Database: STT:bas_...
Liberality: Liberal Query Lang.: fi
Expansion Le...: 1 User name:
Ontologies: Ontologies Onto Lang.: English

Search in Tree:
Ontology: <http://www.info.uta.fi:8880/QuccooFlexServer/ES.xml>
Concept language: Swedish

Result List:

- Vinkkilän Luomutuote <http://www.vinkkilanluomutuote.fi/-15k>
- Vinkkilän Luomutuote <http://www.vinkkilanluomutuote.fi/terveystieto.html> -257k
- Luomutuotteet - Luomut - Elintarvike - Ekotuote - Aromaterapiat ... <http://www.luomukauppa.fi/verkkokauppa/-4k>
- Iso-Markun luomutila <http://www.luomutuote.net/-11k>
- Luomutuote <http://www.luomutuote.fi/-4k>
- MySpace.com - Luomutuote - Espoo, FI - Rock / Punk / Alternative ... <http://profile.myspace.com/index.cfm?fuseaction=user.viewprofile&friendID=2>

Result Page: 1 2 3 4 5 6 7 8 9 10

Ontology Tree (right panel):

- Ontology
- begrepp_av_livsmedelsbransch
 - övervakning_av_livsmedel
 - livsmedel
 - riskfaktor_för_livsmedel
 - djursjukdom
 - giftigt_ämne
 - förgiftning
 - orsak_till_matförgiftning
 - parasit
 - mikrob
 - bakterie
 - virus
 - främmande_ämne
 - handel
 - transport
 - lagsättning
 - förpackningsbeteckningar
 - renhet
 - friskhet

OntoClass (Web interface)

The screenshot shows the OntoClass web interface running in Mozilla Firefox. The main window displays search results for the term "Luomutuote". The search results include:

- Vinkkilän Luomutuote: A description of Vinkkilän Luomutuote as an organic food product.
- Vinkkilän Luomutuote: A link to the official website.
- Luomutuotteet - Luomut - Elintarvikke - Ekotuote - Aromaterapiat ...: A general category for organic products.
- Luomutuote: A link to the Luomutuote website.
- Iso-Markun luomutila: A link to Iso-Markun luomutila's website.
- Luomutuote: A link to the Luomutuote MySpace profile.
- MySpace.com - Luomutuote - Espoo, FI - Rock / Punk / Alternative ...: Another link to the Luomutuote MySpace profile.

The left sidebar shows a tree view of the ontology under the "Ontology" category, specifically the "concepts_of_food_industry" branch. The tree includes categories like food_supervision, food, food_risk_factor, and trade.

The right sidebar shows a tree view of the ontology under the "Ontology" category, specifically the "concepts_of_food_industry" branch. The tree includes categories like food_supervision, food, food_risk_factor, and trade.

At the bottom, there is a navigation bar with links to "Classification", "Search", and "Help".

Discussion

Thank you!

Over to you ... questions?