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

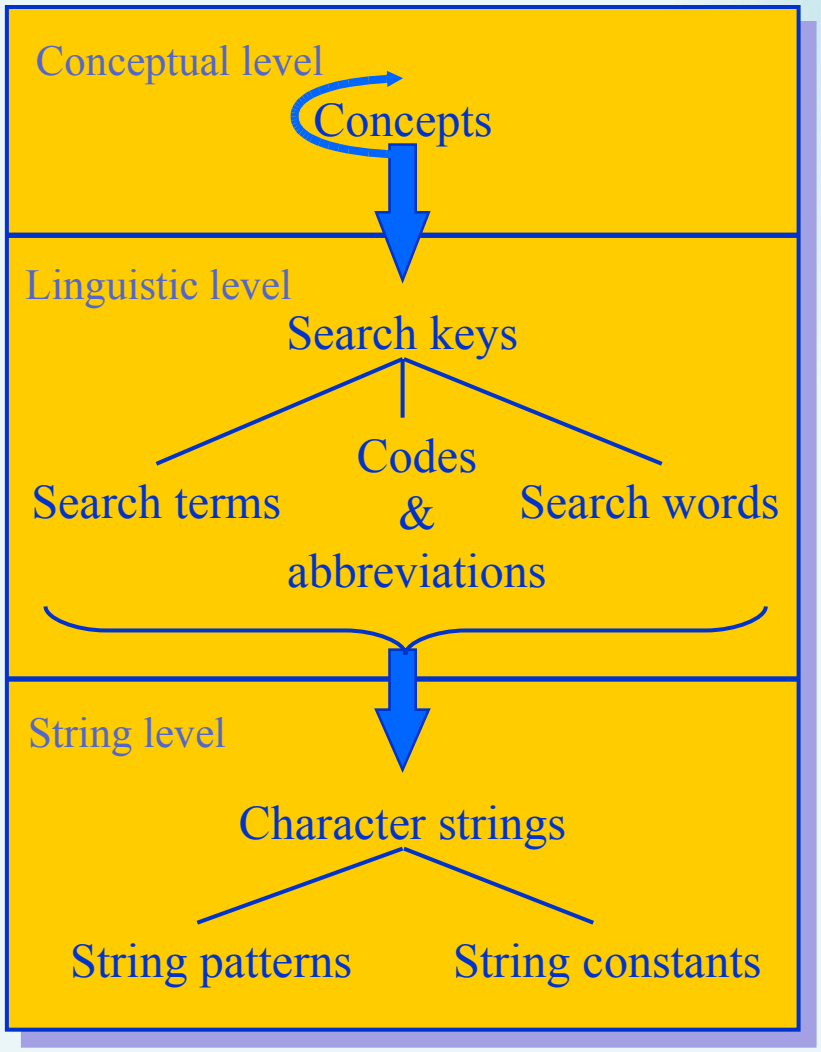
- Index languages -
  - modeling - coverage, viewpoint
  - maintenance - ageing, cost
- Indexing -
  - specificity, exhaustivity, consistency
  - cost - where paid, who pays?
  - The over-specificity the devices created often lead to poor recall and thus they were soon mostly abandoned

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)

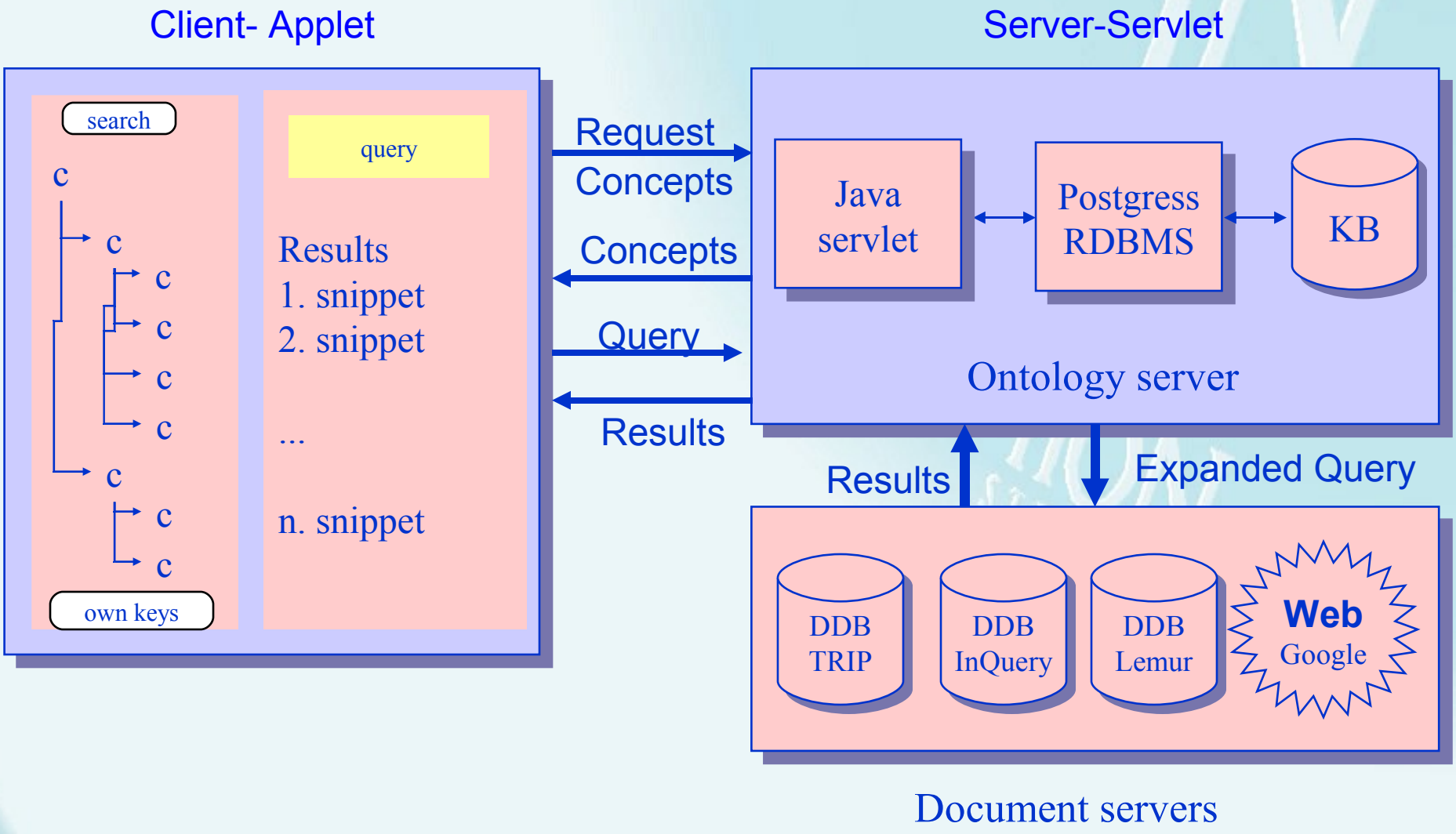
QUCCOO: Principles

- 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

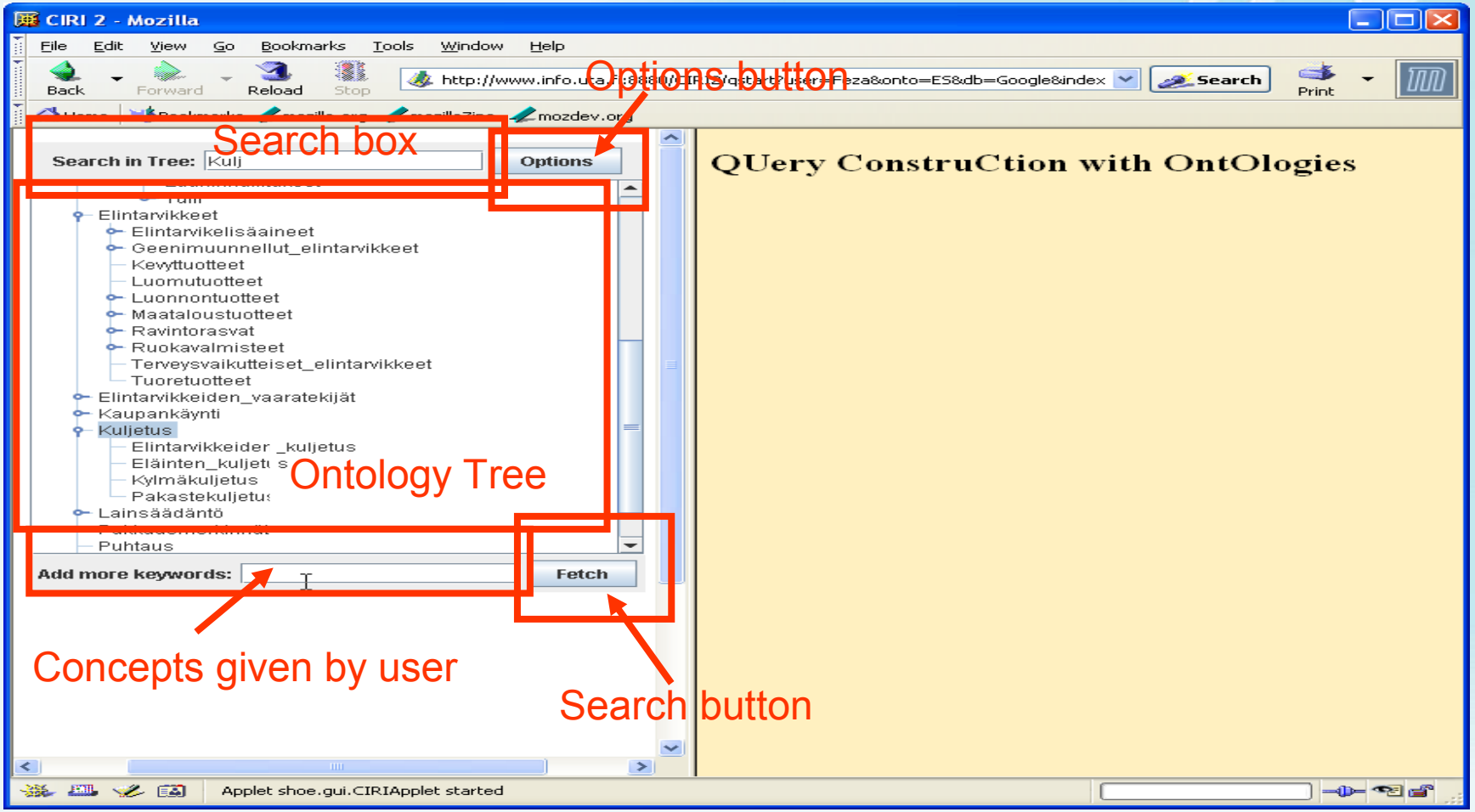
QUCCOO: Status

- 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

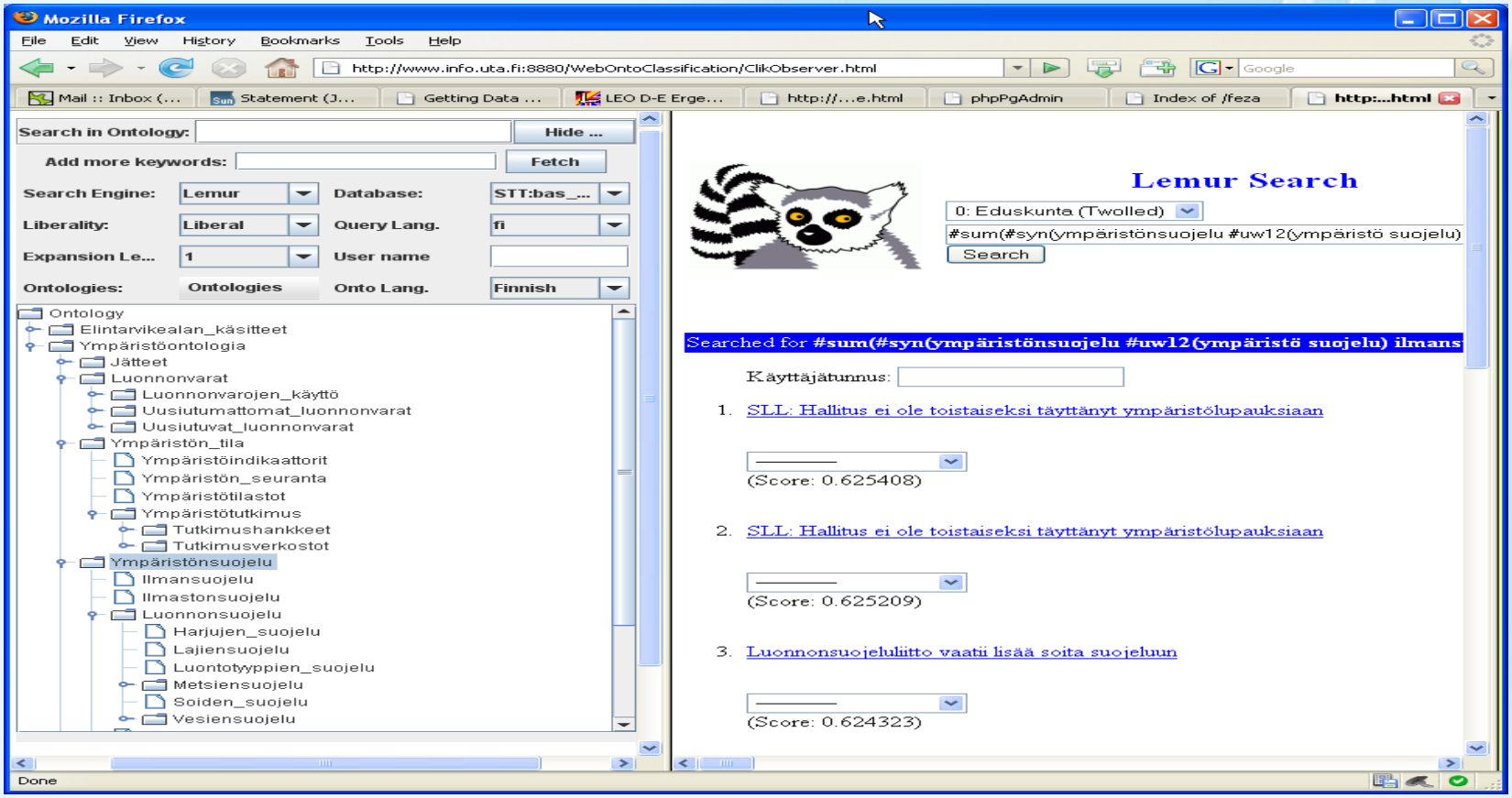
QUCCOO: Architecture



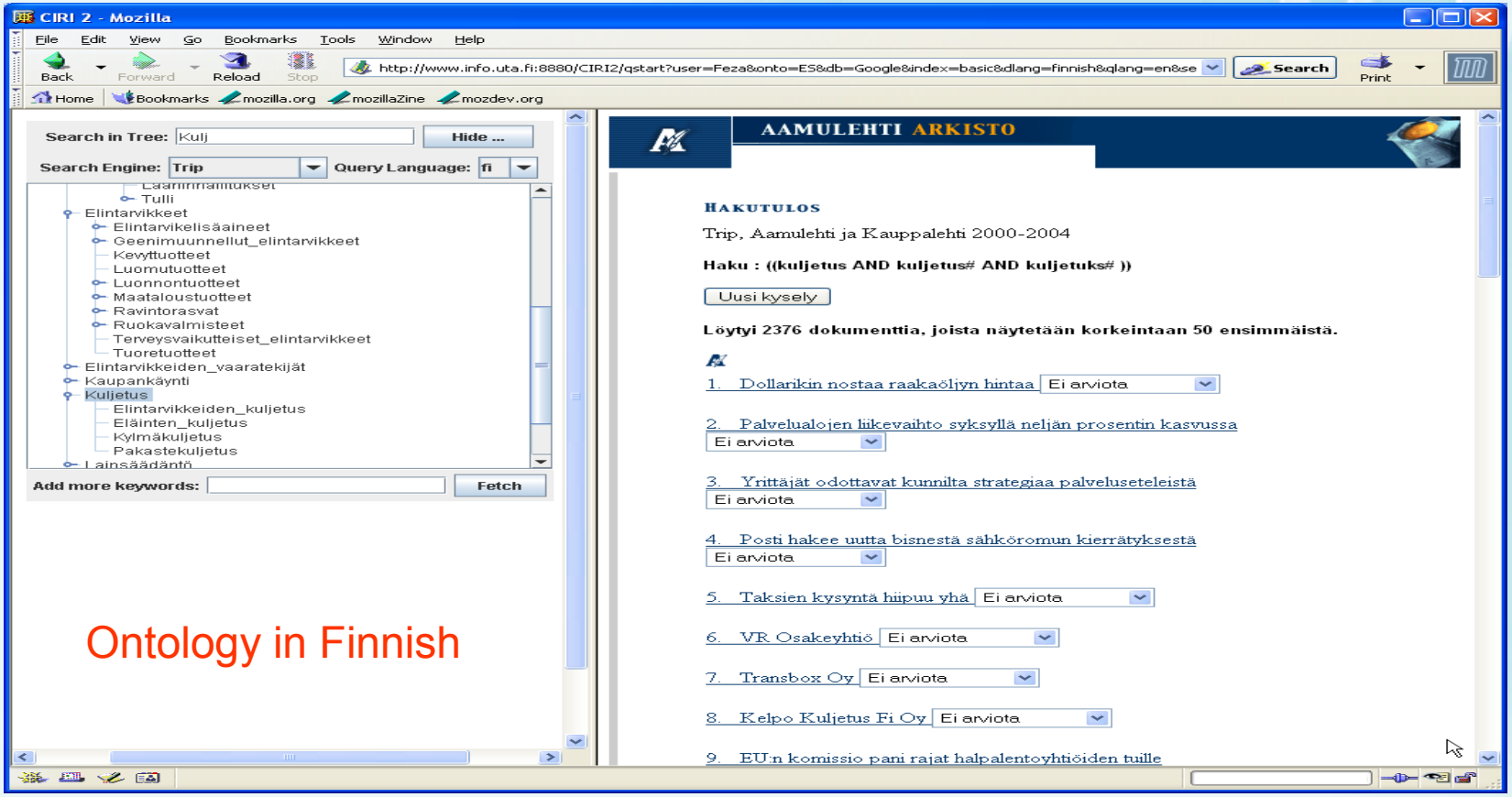
Quccoo - interface



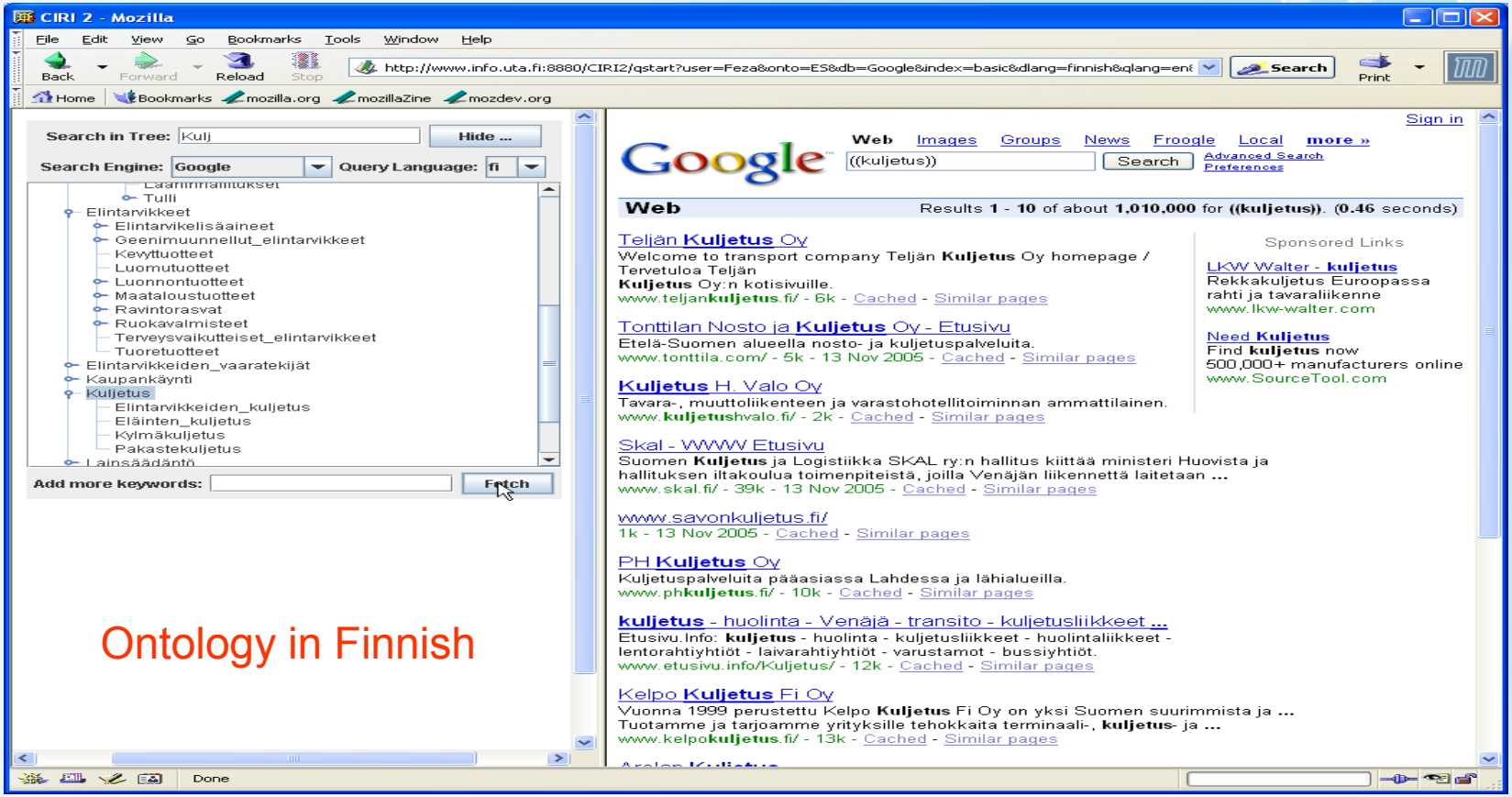
Quccoo - interface



Quccoo - interface

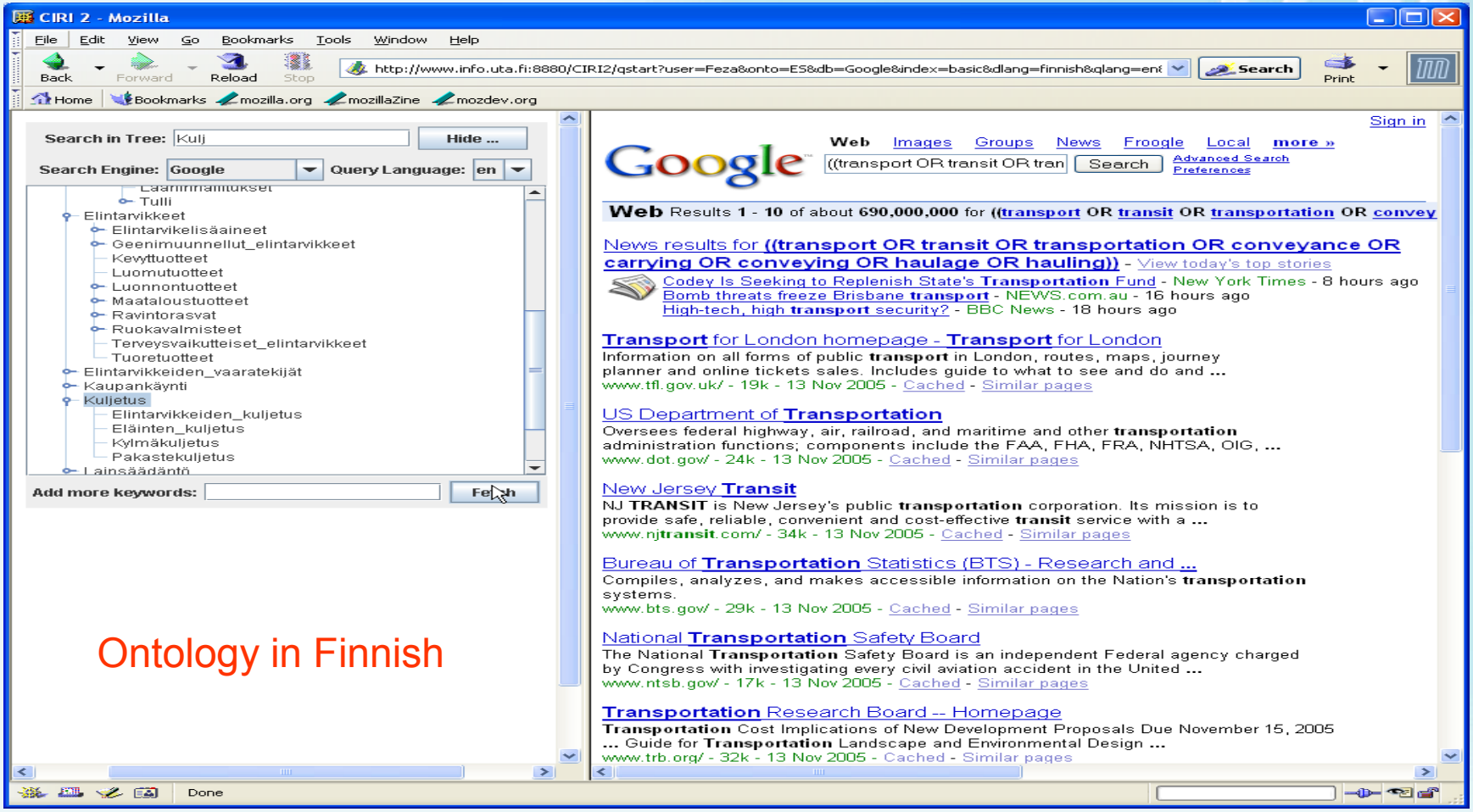


Quccoo - interface

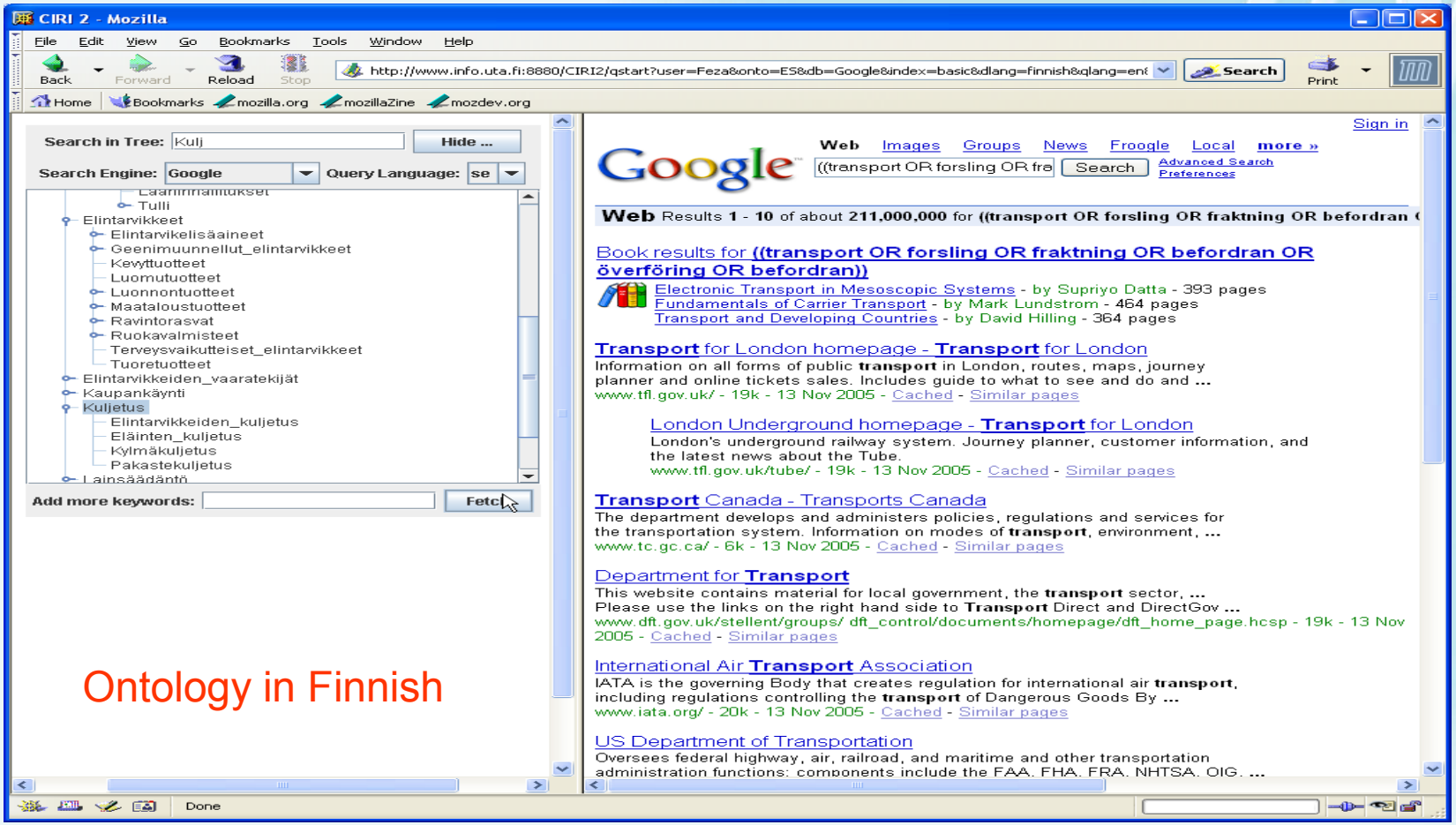


Google Results in Finnish

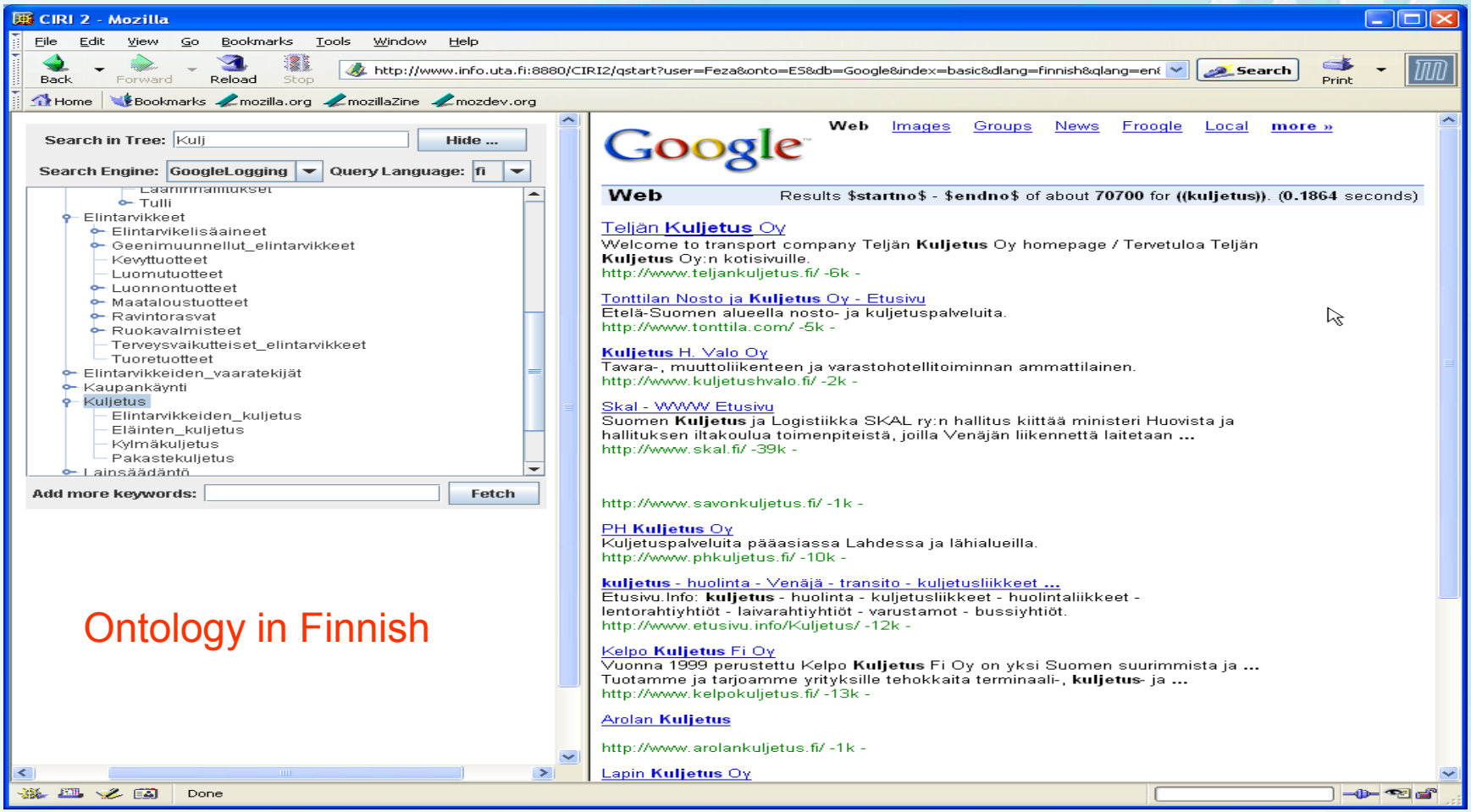
Quccoo - interface



Quccoo - interface



Quccoo - interface



Google Results with Logging Facility

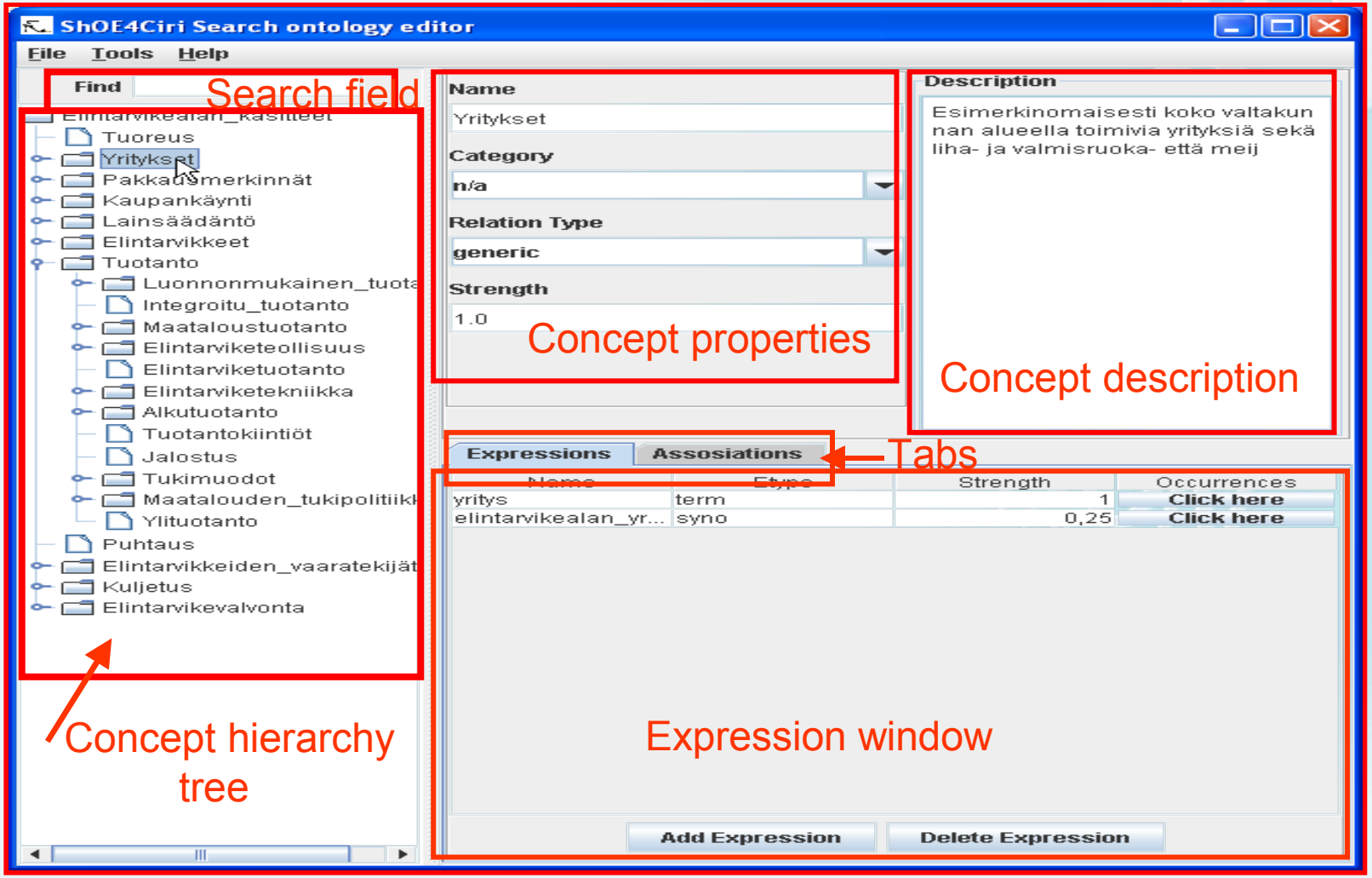
4. ShOE: Creating ontologies

- **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



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

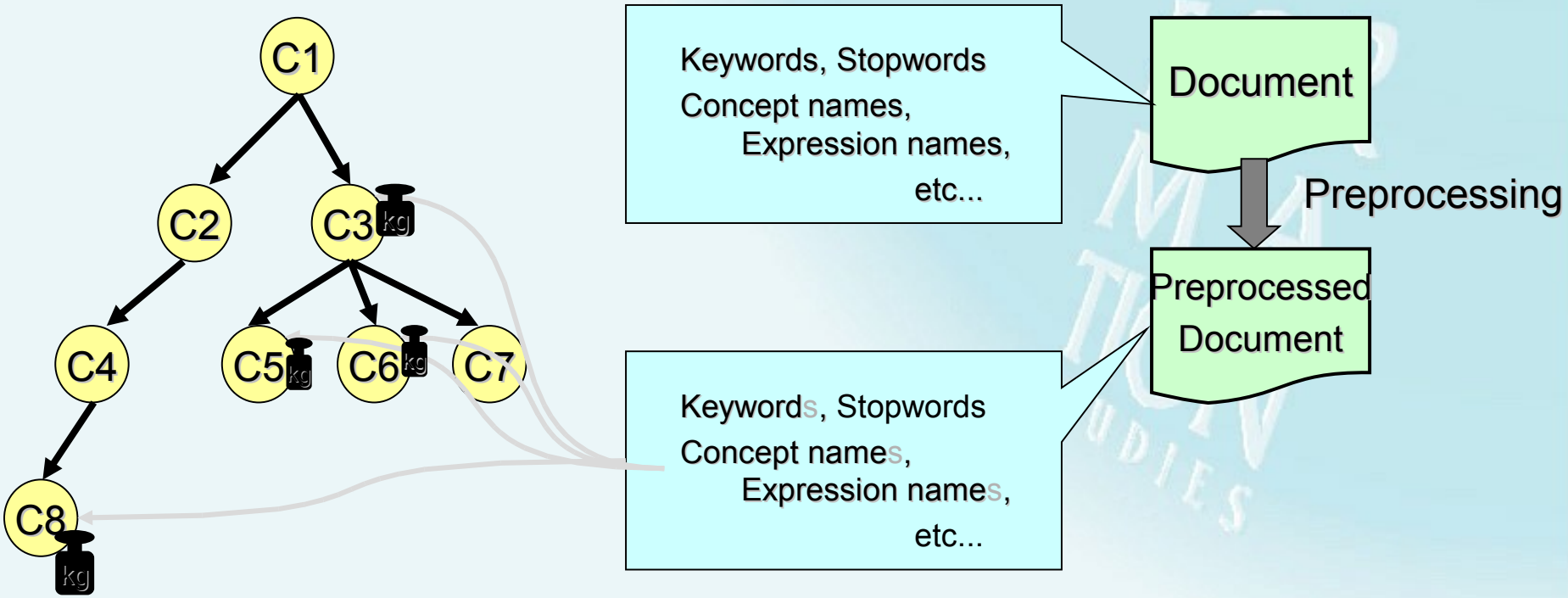
- easy to comprehend and easy to interpret
- based on the visual comprehension faculty of the user

Ontoclass,Ontology based Classification System

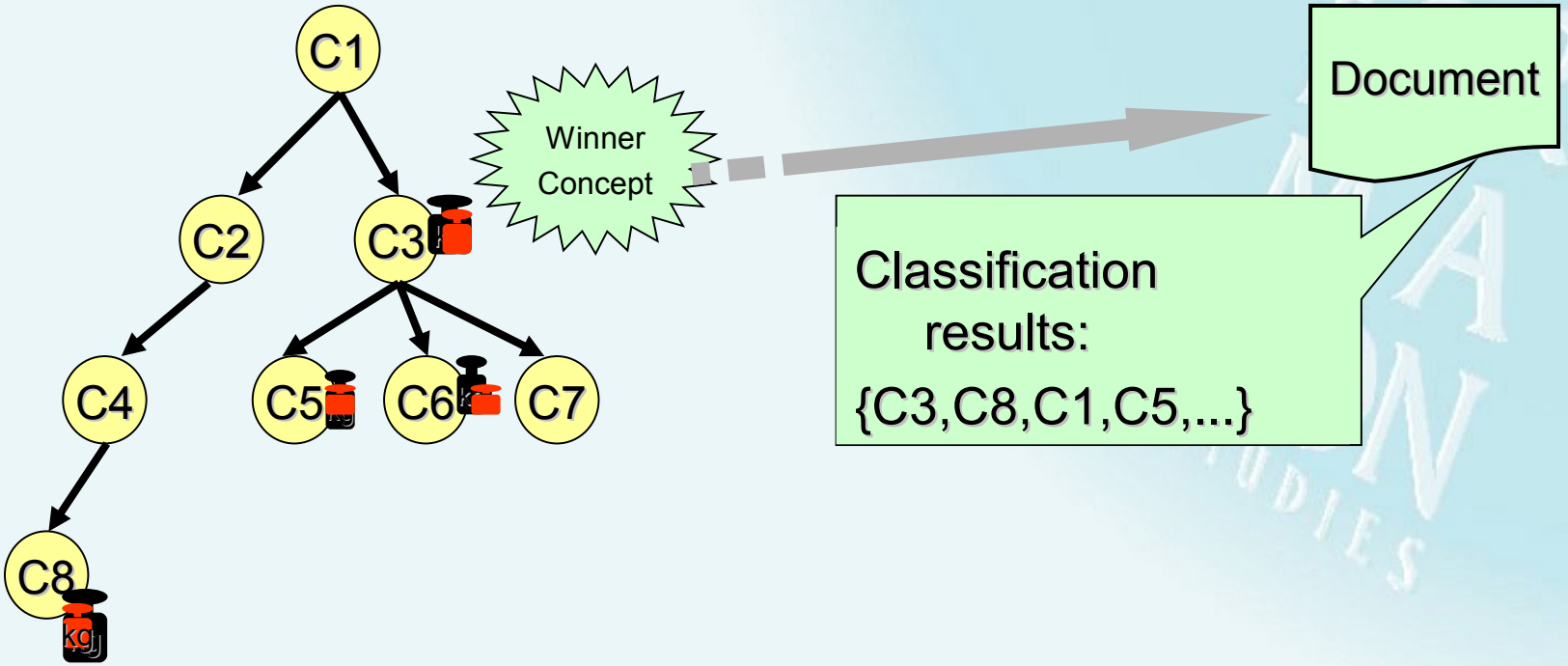
Consists of the following processes:

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

Ontoclass, Ontology based Classification System

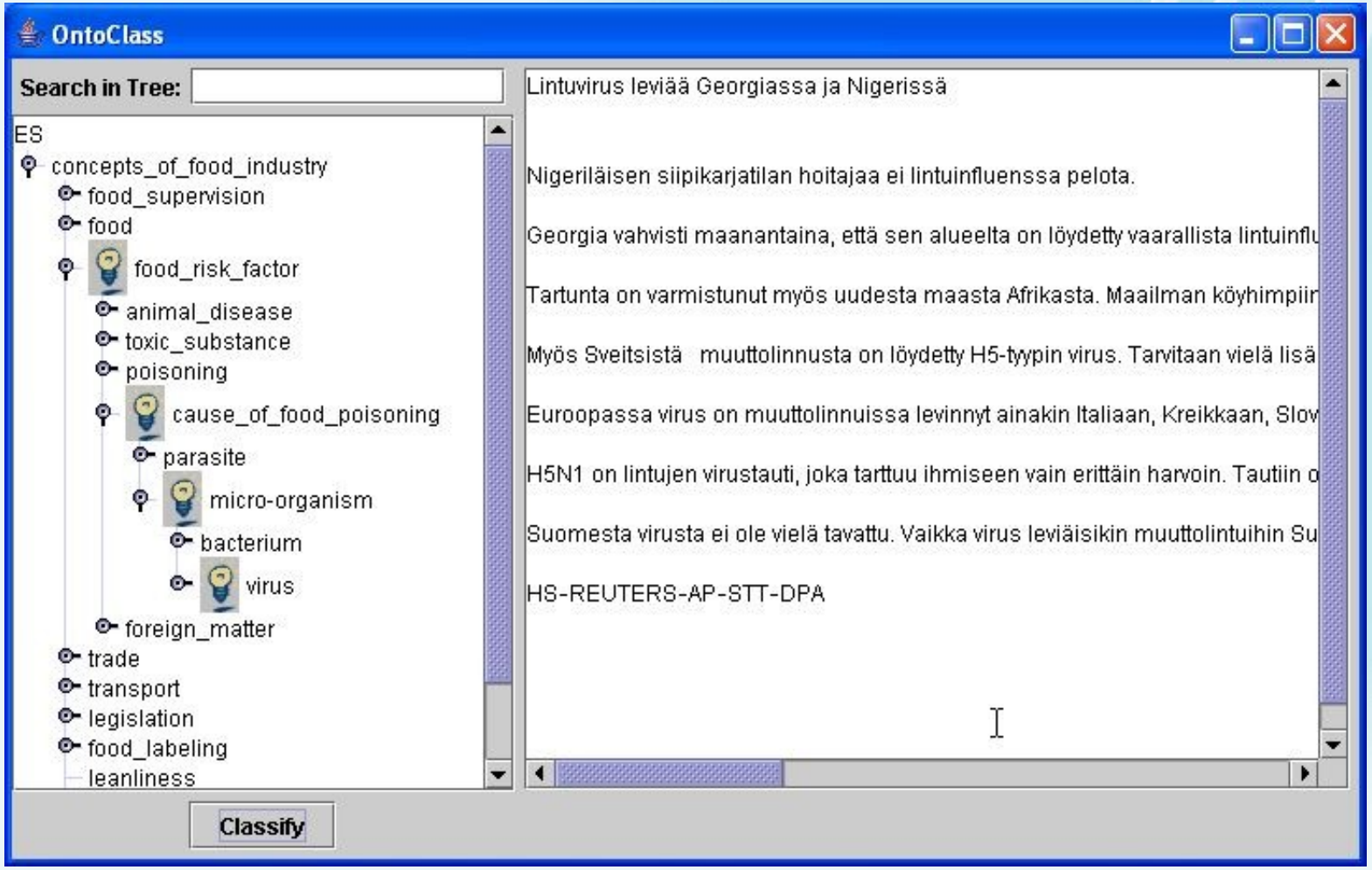


Ontoclass, Ontology based Classification System

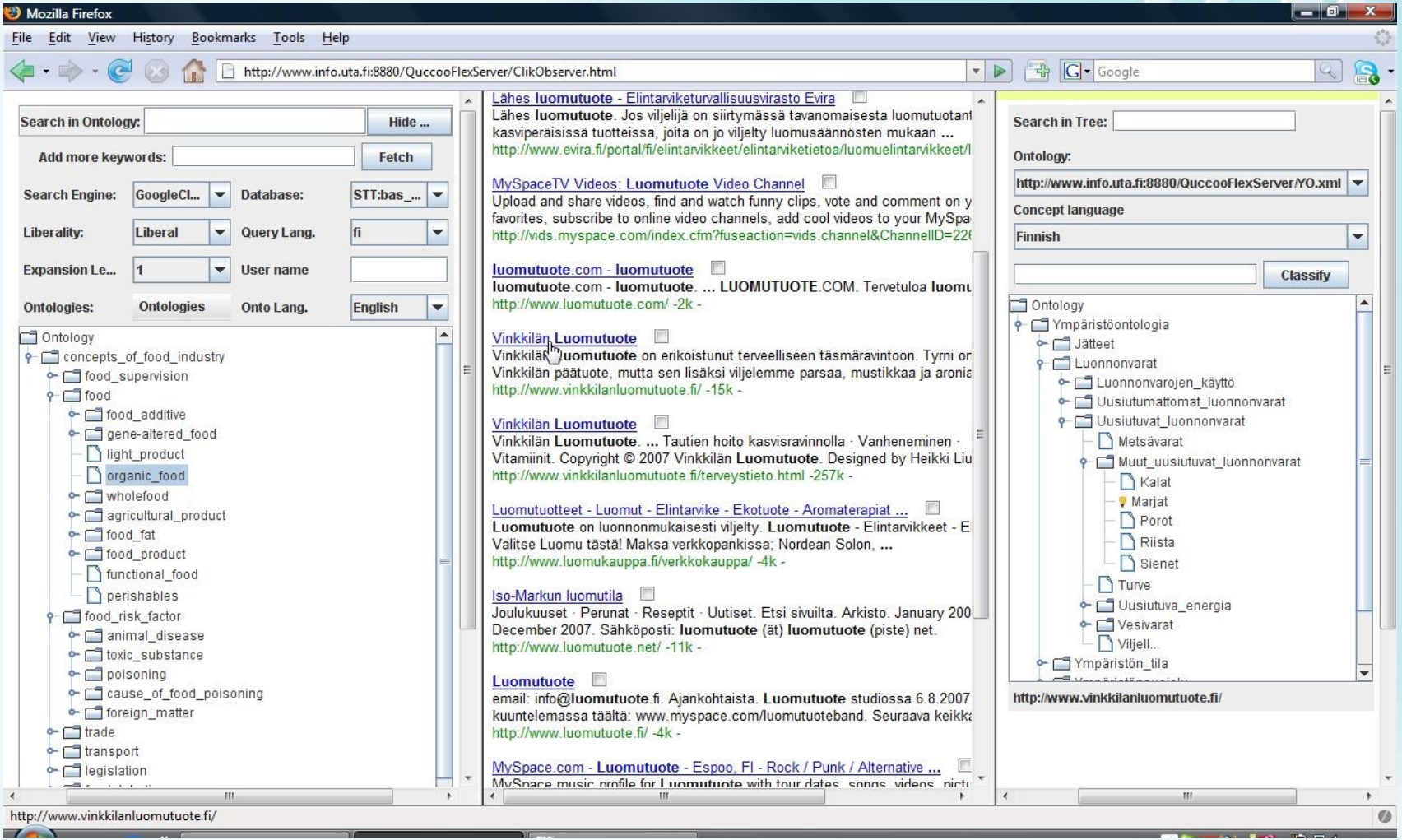


Weight propagation

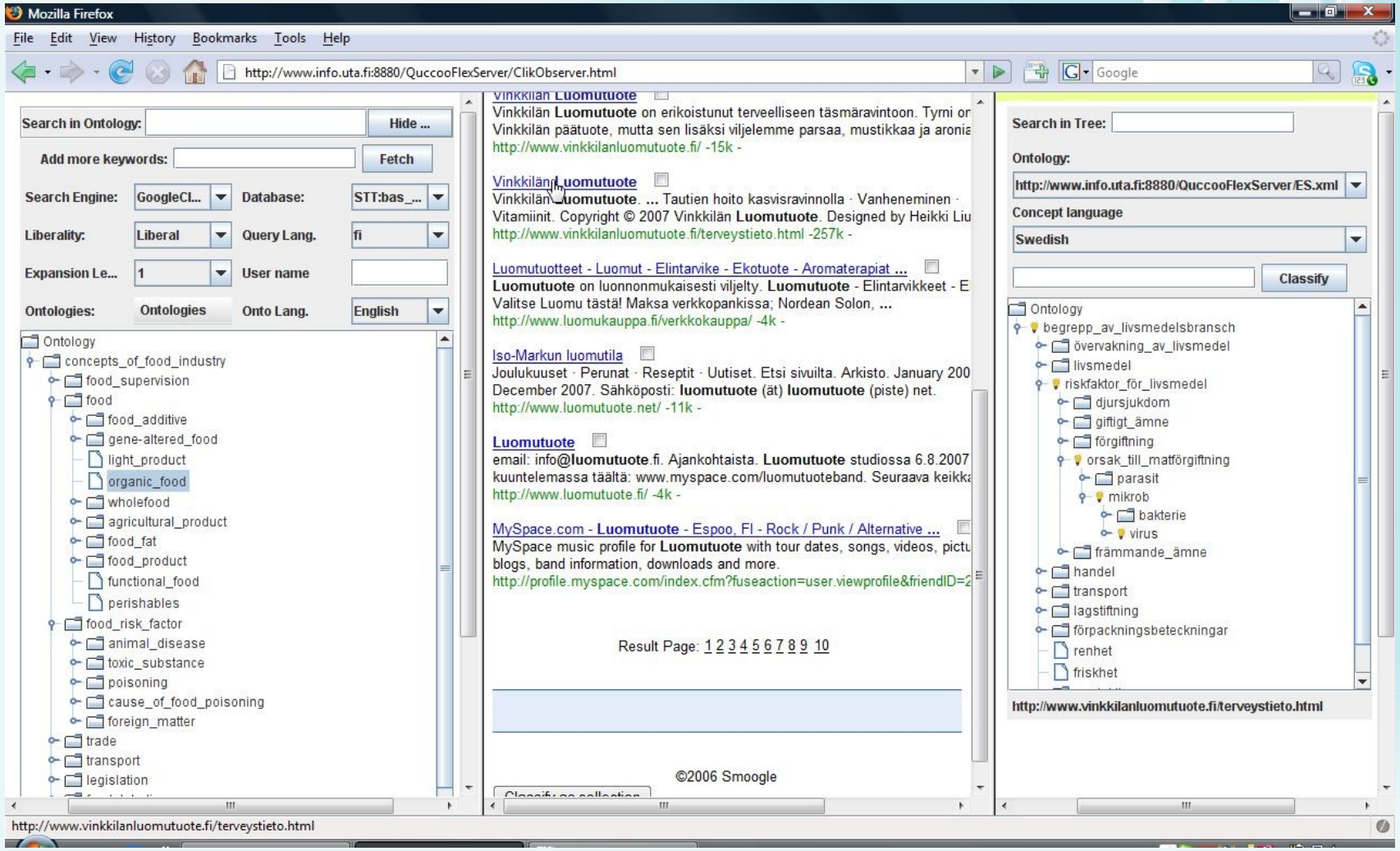
OntoClass
User interface (Standalone version)



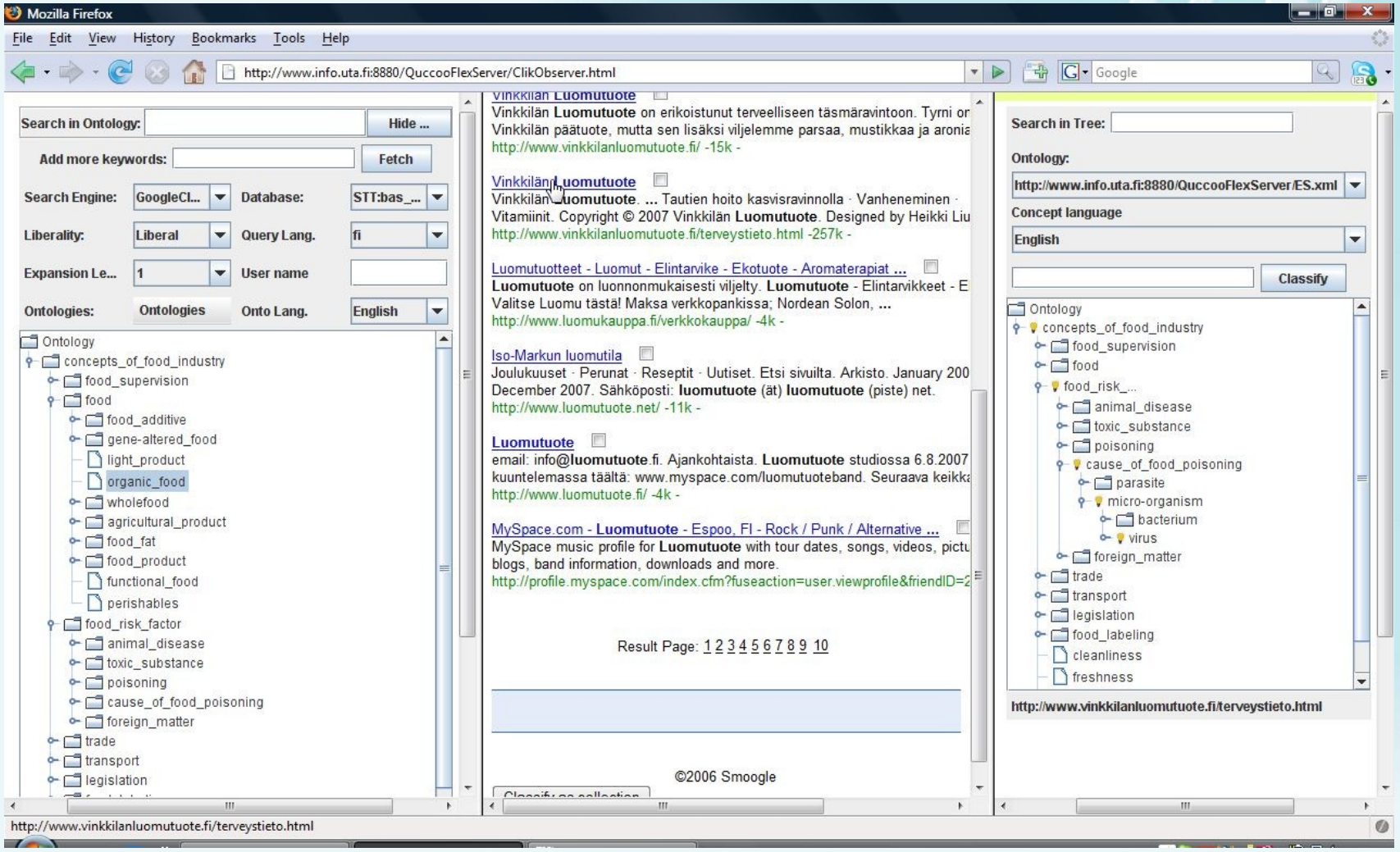
OntoClass (Web interface)



OntoClass (Web interface)



OntoClass (Web interface)



Discussion

Thank you!

Over to you ... questions?