

Second International Seminar
on Subject Access to Information, Helsinki
30th November 2007

QUCCOO

Query Construction with Ontology Ontology-based Search Interface

Feza BASKAYA
Anne KAKKONEN

University of Tampere
Department of Information Studies

Feza Baskaya – Anne Kakkonen - University of Tampere - 2007

1

Outline

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

Feza Baskaya – Anne Kakkonen - University of Tampere - 2007

2

1. Background

- ∅ Vast online information environments
 - § billions of digital documents
 - § many different natural languages
 - § distributed document production and publication: no generally agreed rules
 - § general lack of control in the process
 - § much spam and other unwanted material

Feza Baskaya – Anne Kakkonen - University of Tampere - 2007

3

Background, 2

- ∅ 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

Feza Baskaya – Anne Kakkonen - University of Tampere - 2007

4

2. Ontologies

- ∅ Ontologies model semantics
 - § concepts
 - § rich relationships
 - § support inference
 - § application means resource annotation
 - § closely related to thesauri
- ∅ Belief: 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

Feza Baskaya – Anne Kakkonen - University of Tampere - 2007

5

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

Feza Baskaya – Anne Kakkonen - University of Tampere - 2007

6

Any Room for Ontologies?

- ∅ Should one thus discard ontologies?
 - § or other vocabulary control tools?
- ∅ In practice, realism tells us that
 - § there will never be a comprehensive & up-to-date ontology – cf. UDC, which *had* large development community support
 - § no one will annotate for free, for ever & consistently
 - § no one can do that exhaustively and from many viewpoints emerging, e.g., in future
 - § in fact, less than 0.3% of web pages had Dublin Core metadata (Rasmussen 2003)
- ∅ There is *no alternative* to searching unannotated collections
 - § automatic annotation does not solve the problem - if one aims at the good semantics required in the Semantic Web

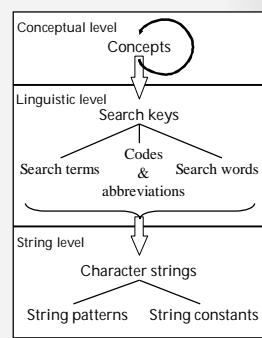
3. Searching Unannotated Collections through Ontologies

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

Searching through Ontologies

- ∅ Need to solve the vocabulary problem from concepts to textual expressions
 - § three layers:
 - ü Concepts - for user interaction
 - ü Expressions - for system use
 - ü Strings to match - for system use
- ∅ Need to provide a handy concept browser and query constructor

Three levels



∅ Forest industry

- § forest industry
- § paper industry
- § saw mill
- § ...
- § pl(saw, mill)
- § al(industry)
- § pl(paperi, tehdas)

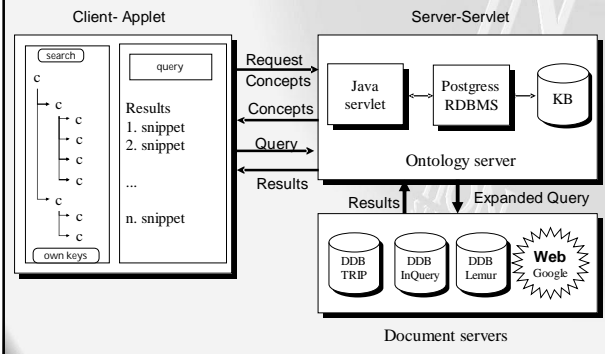
QUCCOO: Principles

- ∅ 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

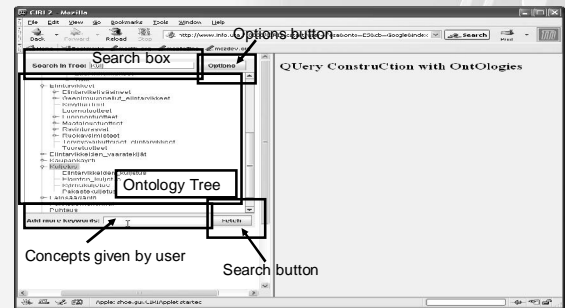
QUCCOO: status

- ∅ 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



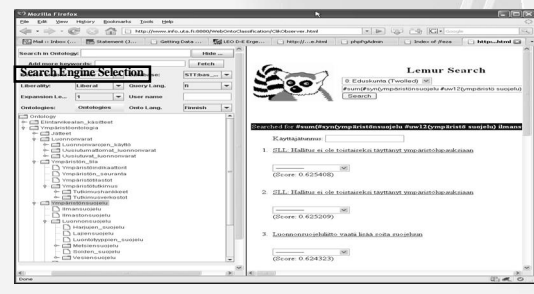
QUCCOO - interface



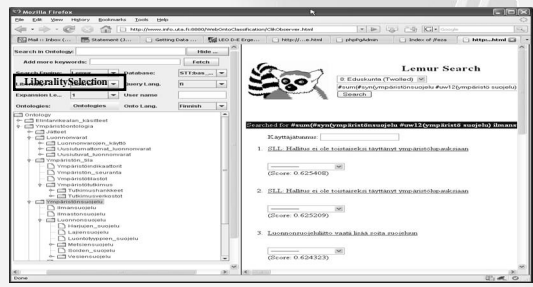
QUCCOO - interface



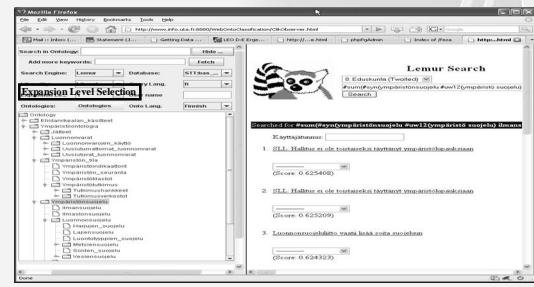
QUCCOO - interface



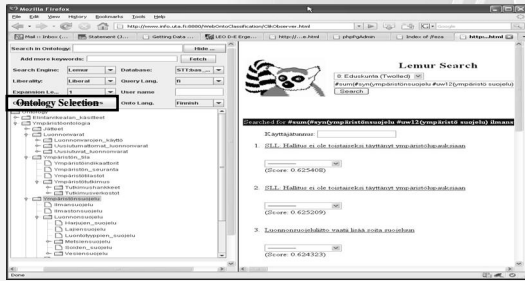
QUCCOO - interface



QUCCOO - interface

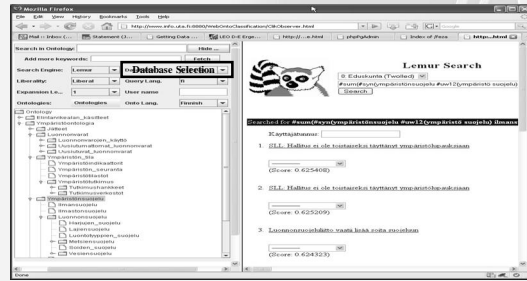


QUCCOO - interface



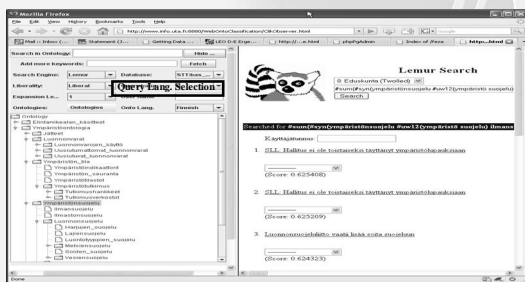
Feza Baskaya – Anne Kakkonen - University of Tampere - 2007

QUCCOO - interface



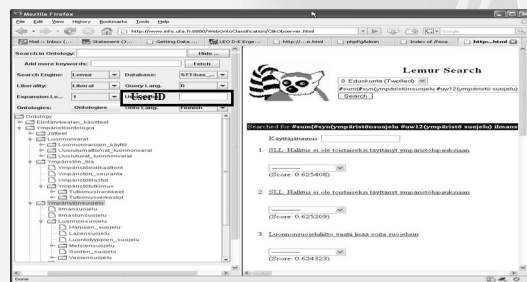
Feza Baskaya – Anne Kakkonen - University of Tampere - 2007

QUCCOO - interface



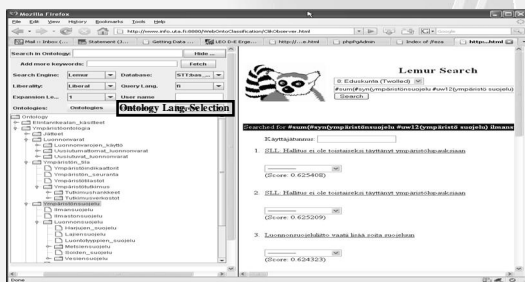
Feza Baskaya – Anne Kakkonen - University of Tampere - 2007

QUCCOO - interface



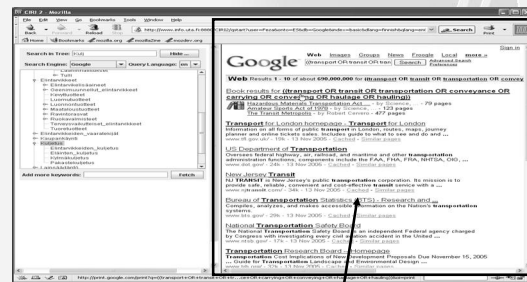
Feza Baskaya – Anne Kakkonen - University of Tampere - 2007

QUCCOO - interface



Feza Baskaya – Anne Kakkonen - University of Tampere - 2007

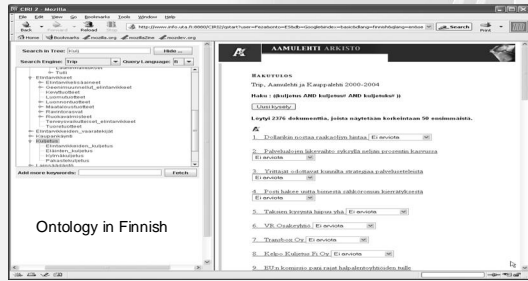
QUCCOO - interface



Query result page

Feza Baskaya – Anne Kakkonen - University of Tampere - 2007

QUCCOO - interface



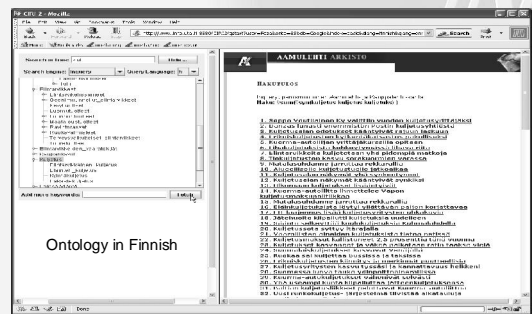
Ontology in Finnish

Trip Database Engine results

Feza Baskaya - Anne Kakkonen - University of Tampere - 2007

25

QUCCOO - interface



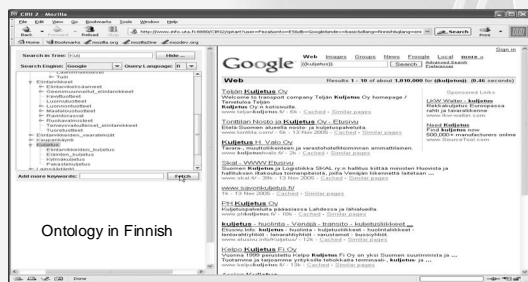
Ontology in Finnish

Inquiry Database Engine Results

Feza Baskaya - Anne Kakkonen - University of Tampere - 2007

26

QUCCOO - interface



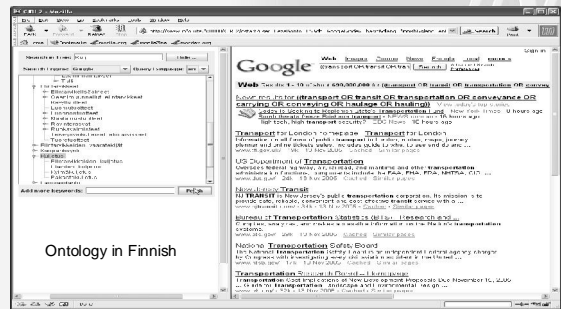
Ontology in Finnish

Google Results in Finnish

Feza Baskaya - Anne Kakkonen - University of Tampere - 2007

27

QUCCOO - interface



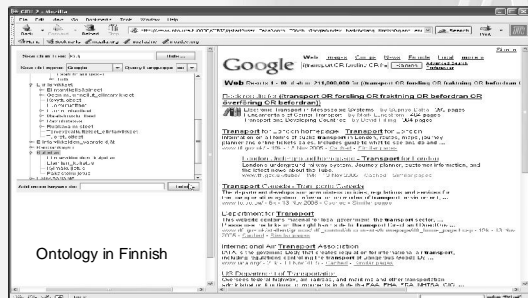
Ontology in Finnish

Google Results in English

Feza Baskaya - Anne Kakkonen - University of Tampere - 2007

28

QUCCOO - interface



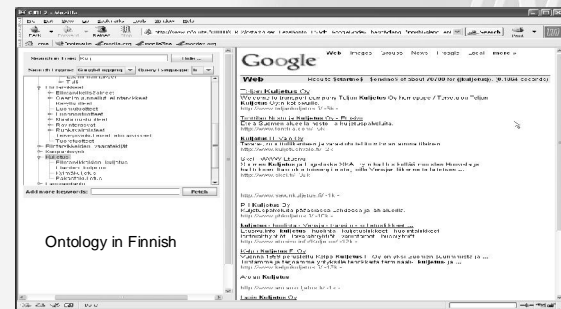
Ontology in Finnish

Google Results in Swedish

Feza Baskaya - Anne Kakkonen - University of Tampere - 2007

29

QUCCOO - interface



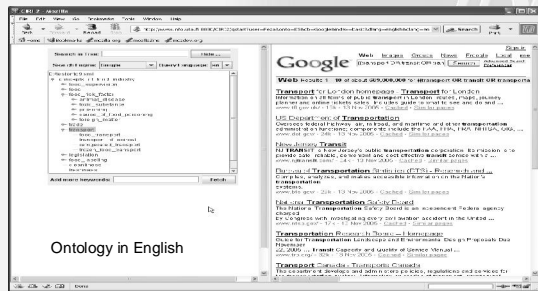
Ontology in Finnish

Google Results with Logging Facility

Feza Baskaya - Anne Kakkonen - University of Tampere - 2007

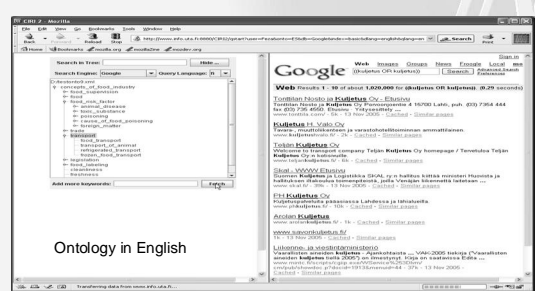
30

QUCCOO - interface



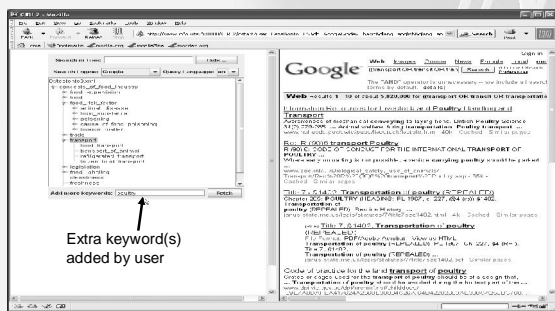
Google Results in English

QUCCOO - interface



Google Results in Finnish

QUCCOO - interface



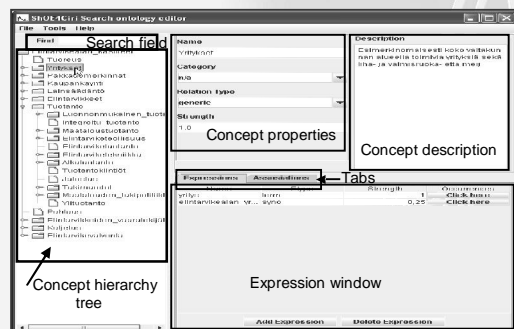
4. ShOE: Creating ontologies

- ∅ 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)

Discussion

Thank you!

Over to you ... questions?