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

QUCCOO
Query Construction with Ontology-based Search Interface

Feza BASKAYA
Anne KAKKONEN

University of Tampere
Department of Information Studies
Outline

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

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
    - 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
Searching through Ontologies

Need to solve the vocabulary problem from concepts to textual expressions

- three layers:
  - Concepts - for user interaction
  - Expressions - for system use
  - Strings to match - for system use

Need to provide a handy concept browser and query constructor
Three levels

- **Conceptual level**: Concepts
- **Linguistic level**: Search keys
  - Search terms
  - Codes & Search words abbreviations
- **String level**: Character strings
  - String patterns
  - String constants

- 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

```
search

query

Results
1. snippet
2. snippet
...
n. snippet

own keys
```

Server-Servlet

```
Java servlet

Postgres RDBMS

KB

Ontology server

Request Concepts

Concepts

Query

Results

Expanded Query

Document servers

```

Request

Results

KB

```

DDB TRIP

DDB InQuery

DDB Lemur

Web

Google
Quccoo - interface

Search box

Options button

Ontology Tree

Concepts given by user

Search button
Quccoo - interface
Quccoo - interface

Search Engine Selection

Lemur Search

1. SLL: Hallitus ei ole toistaiseksi tavoitettu ympäristösuojelua
   (Score: 0.625408)

2. SLL: Hallitus ei ole toistaiseksi tavoitettu ympäristösuojelua
   (Score: 0.623209)

3. Luonnonsuojeluotto vaati lieän sidoa suojelua
   (Score: 0.624323)
QuccoO - interface
Quccoo - interface

Expansion Level Selection

Lemur Search

Search for \#sum(#syn(ympäristönsuojelu #uw12(ympäristö suojelu))

Käyttäjänimi: [Input]

1. SLL Häälne ei ole toistamiseksi tarvittavat ympäristönsuojelut
   (Score: 0.625488)

2. SLL Häälne ei ole toistamiseksi tarvittavat ympäristönsuojelut
   (Score: 0.625209)

3. Lounasoravirhe vaati lisää soita suojelun
   (Score: 0.624323)
**Quccoo - interface**

Database Selection

Lemur Search

1. SLL: Hallitus eik ole toistaiseksi syrhtnyt ympyrystystsuojelula

(Score: 0.625408)

2. SLL: Hallitus eik ole toistaiseksi syrhtnyt ympyrystystsuojelula

(Score: 0.623209)

3. Luonnonsuojelutoon vaati lis kota rota suojelula

(Score: 0.624323)
Quccoo - interface
Quccoo - interface
Quccoo - interface
Quccoo - interface

Query result page
Quccoo - interface

Ontology in Finnish

Trip Database Engine results
Quccoo - interface

Ontology in Finnish

Inquiry Database Engine Results
Quccoo - interface

Ontology in Finnish

Google Results in Finnish
Quccoo - interface

Ontology in Finnish

Google Results in English
Quccoo - interface

Ontology in Finnish

Google Results in Swedish
Quocco - interface

Ontology in Finnish

Google Results with Logging Facility
Quccoo - interface

Ontology in English

Google Results in English
Quccoo - interface

Extra keyword(s) added by user
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
- Concept description
- Tabs
- Expression window

Feza Baskaya – Anne Kakkonen - University of Tampere - 2007
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?