Compass

Improved Findability Through Semantic Search

www.ovitas.no
About us

- We are Eszter & Rolf
- Ovitas AS - information management since 1996 involving SGML/XML based technologies
- Engaged in the development of ontology based search and navigation
Searching approaches

- Traditional search is rarely topic centred
  - Full text search engines, e.g. Lucene are getting better, but they don’t say anything about how topics are related to each other
  - Yahoo - directory search *
    “Directory is an online guide to the World Wide Web. It is catalog of sites created by a staff of editors who visit and evaluate web sites, and then organize them into subject-based categories and sub-categories.”

- We were searching for a better approach...
INSIDE YAHOO!

Latest Health information on Yahoo! Health

CATEGORIES (What's This?)

- Alternative Medicine (562)
- Chats and Forums (44)
- Children's Health (216)
- Conferences (17)
- Consumer Products and Services@
- Death and Dying@
- Dental Health (77)
- Disabilities@
- Diseases and Conditions (10944) NEW!
- Education (52)
- Emergency Services (530)
- Medical Geography@
- Medicine (4757)
- Men's Health (55)
- Mental Health (802)
- Midwifery (44)
- News and Media (221)
- Nursing (403)
- Nutrition (409)
- Organizations (22)
- Pet Health@
- Pharmacy (1855)
Searching approaches

- Traditional search is rarely topic centred
  - Full text search engines, e.g. Lucene are getting better, but they don’t say anything about how topics are related to each other
  - Yahoo - directory search *
    “Directory is an online guide to the World Wide Web. It is catalog of sites created by a staff of editors who visit and evaluate web sites, and then organize them into subject-based categories and sub-categories.”

- We were searching for a better approach…
What’s not working?

- Too many hits
- Too few hits
- No hits
- Reasonable number of hits, but not the relevant ones
- Difficult to find the right search words
- Hits are not categorized
Can we fix the problem?

- No need to reinvent the wheel!
- Existing components:
  - Text analyzing tools
  - Full text search engine
  - Topic Maps engine
- Combine them in a new way...
Basic ideas of Compass

- Use a knowledge model (e.g. Topic Maps) to describe your domain
- Assign some numeric value that defines the degree of relevance between the topic types in the model
- Enhance the full text search according to the model, retrieving an expanded set of search terms
- Set a threshold for relevance
- Return weighted search results
What does using a knowledge model really mean?
**Known symbols**

### Land forms
- Terrace
- Spur
- Re-entrant
- Earth bank
- Quarry
- Earth wall
- Erosion gully
- Small erosion gully
- Hill
- Knoll
- Saddle
- Depression
- Small depression
- Pit
- Broken ground
- Ant hill

### Water and marsh
- Lake
- Pond
- Waterhole
- River, Stream, Watercourse
- Minor water channel, Ditch
- Narrow marsh
- Marsh
- Firm ground in marsh
- Well
- Spring
- Water tank, Water trough

### Vegetation
- Open land
- Semi-open land
- Forest corner
- Clearing
- Thicket
- Linear thicket
- Vegetation boundary
- Copse
- Distinctive tree
- Tree stump, Root stock

### Man-made features
- Road
- Track/Path
- Ride
- Bridge
- Power line
- Power line pylon
- Tunnel
- Stone wall
- Fence
- Crossing point
- Building
- Paved area
- Ruin
- Pipeline
- Tower
- Shooting platform
- Boundary stone, Cairn
- Fodder rack
- Platform
- Monument or Statue
- Building pass through
- Stairway

### Special features
- Special item
- Special item
Known symbols put in relation to each other
The metaphor

- The map is a drawing of the real life terrain
- The topic map is a ”drawing” of the domain we want to talk about
- Maybe old news, but: We offer you the compass!
Example topic map

Ovitas

hasProduct

type

Compass

hasEmployee

type

Christopher
Example topic map

Ovitas

hasProduct

hasEmployee

Compass

Christopher

CW=0.8

type

type

CW=0.7
Architecture

- Service oriented architecture
- Web service interface
- Currently on the .NET platform
- Integrated components:
  - *TMCore* Topic Maps engine by *NetworkedPlanet*
  - *Apache Lucene*: full text search engine
Architecture diagram
Compass orchestrator

- Guides the processes of the search:
  - User enters a search term
  - Search for term in the topic map
  - Look in the map for relevant/related topics
  - Send all these terms off to a fulltext search
  - Calculate relevance (based on the combination of CompassWeight and Lucene weights) and prepare the result list as an XML instance
  - Render XML results
Search Results for:
Compass

You got a total of 55 hits for Compass

Grouped hits
- Compass Hits: 6
- Ovitas AS Hits: 33
- Christopher Kolstad Hits: 1
- Eszter Horvati Hits: 1
- Information Retrieval Hits: 3
- C# Hits: 6
- Topic Maps Hits: 31

10 Most relevant hits
- **Compass.pdf**
  that combine the right search and navigation methods. COMPASS Intelligent Information Retrieval COMPASS - Information Retrieval Solution intelligent information retrieval enables you to find ... for information retrieval. Compass can be used in searching through vast amounts of corporate
  http://www.ovitas.no/external/display/Compass.pdf

- **Ovitas AS**
  Horvati er ansatt som informasjonsarkitekt. Christopher Kolstad er ansatt som Systemutvikler ... Nyheter Nyheter Ovit... konferansen i Oslo den 29. mars. Læs mer om Compass her... SAS velger Ovit... lansering
  http://www.ovitas.no/external/news.do

Related topics
- Company
- Ovitas AS
- Person
- Barnebas Davoti
- Christopher Kolstad
- Eszter Horvati
- Skill
- C#
- Lucene
- Topic Maps
- Solution
- Information Retrieval

Search term

Hits in the fulltext grouped by the related topics

Topic Map expansion

Relevant documents ranked by the weighting result
Synonym search

Search Results for

Emnekart

You got a total of 38 hits for Topic Maps Emnekart

Grouped hits

- Topic Maps Hits: 31
- Emnekart Hits: 10
- Christopher Kolstad Hits: 1
- Graham Moore Hits: 4
- Kal Ahmed Hits: 6
- Compass Hits: 6
- TMC0re05 Hits: 21
- Eszter Horvati Hits: 1
- Technologies Hits: 2

10 Most relevant hits

Ovitas AS

A Topic Maps engine (emnekart) TMC0re05 fra Networked Planet gir utviklere muligheten til å dra nytte av alle fordelene som emnekart (topic maps) gir i sine applikasjoner. Emnekartmotoren inneholder et... løsning. Semantiske søk basert på emnekart (Topic Maps). Organiser og grupper søkeresultatet i

http://www.ovitas.no/external/technology.do?id=38

NetworkedPlanet - Networked Information, Networked Knowledge

Kal Ahmed and Graham Moore of NetworkedPlanet have contributed an article describing topic maps... with topic maps. CRF Format Technologies Network/Planet.
Creating/maintaining the model

- An MS Excel plug-in serves as the topic map editor
- Can be put under version control
- Import the model into the topic map engine: one click only
- In addition: Web service XTM import
Navigation

- Navigation through the associations between topics
- Navigation by search
User configurations

- What pages to index
- What topic map to use
- The number of hops to perform
- The threshold for relevance

```
Ovitas
  └── hasProduct
    └── type
        └── Compass
          └── hasEmployee
            └── type
                └── Christopher
```

\[ CW = 0.8 \]
\[ CW = 0.7 \]
\[ CW = 0.56 \]
Compass works well when...

- We have a web site with a lot of text
- Which is unstructured (no markup, no tags)
- A controlled domain (we know what the domain of discourse is)
Examples
Søk på Morfin Ga: 93 treff.

Eksakt treff

- Morfin «Nycomed Pharma» Fra: Preparater
- Morfin Epidural «Nycomed Pharma» Fra: Preparater
- Morfin-Skopolamin «Nycomed Pharma» Fra: Preparater
- N02A A01 Morfin Fra: Interaksjoner

Grupperte treff

- Preparater (32)
- ATC-koder (34)
- Firmaer (7)
- Antidoter (2)
- Gifter (1)
- Formelregister (1)
- Graviditet og amming (2)
- Interaksjoner (10)
- Substans (1)
- Diverse dokumenter (2)
- Endrede tekster (1)
Søk på Keflex Ga: 10 treff.

Eksakt treff

Keflex «EuroCept» Fra: Preparater

Grupperte treff

- Preparater (4)
- ATC-koder (2)
- Firmaer (1)
- Endrede tekster (3)

Relatert Informasjon

Firma
- EuroCept (0,4)

Preparat
- Keflex «EuroCept» (1)
- Cefalexin «Merck NM» (0,49)

Substans
- J01D B01 Cefalexin (0,7)
Søk på Keflex Ga: 9 treff.

Eksakt treff

Keflex «EuroCept» Fra: Preparater

Grupperte treff

- Preparater (3)
- ATC-koder (2)
- Firmaer (1)
- Endrede tekster (3)
Søk på Kvalme Ga: 113 treff.

Grupperte treff

- Preparater (87)
  - Postafen «UCB»
  - Ondansetron Nycomed «Nycomed Pharma»
  - Alimta «Lilly»
  - Antineuralgica «Alpharma»
  - Marzine «Pfizer»
  - Zofran «GlaxoSmithKline»
  - Stemetil «Aventis Pharma»
  - Sevorane «Abbott»
  - Morfin Epidural «Nycomed Pharma»
  - Carboplatin «Mayne»
  - Phenergan «Aventis Pharma»
  - Phos-Ex «Vitaline»
  - Robinul-Neostigmin «Meda»
  - Kull «Abigo»
  - Primperan «Sanofi-Synthelabo»
  - Cefalexin «Merck NM»
  - Diamox «Goldshield»
  - Diamox Depot «Goldshield»
  - Flutamid «Orion»
  - Ketorax «Pfizer»
  - Travello «Pfizer»
  - Trimetoprim «Orion»
  - Loratadin «Merck NM»
  - Atarax «UCB»
  - Fungizone «Bristol-Myers Squibb»
  - Hexabrix «Guerbet»
  - Kaleorid «LEO»
  - Megace «Bristol-Myers Squibb»
  - Niferex «SCHWARZ PHARMA»
  - Kodein «Nycomed Pharma»

Relatert Informasjon

Substans
A04 Antiemetika (0,7)
Summary

- Semantic search based on Topic Maps
- Get relevant hits within a meaningful context
- Index and search any document formats
- Organize information in a topic-oriented manner
- Link to relevant information without touching the data content
- Tools for maintaining/evolving the model
- Conceptual navigation through search enabled by Topic Maps