WebExplorer
A TOOL FOR ONTOLOGY-BASED INFORMATION EXPLORATION

Feza Baskaya, Anne Keskimaa,

Jaana Kekäläinen, Kalervo Järvelin

University of Tampere, Finland

TRIM
TAMPERE RESEARCH CENTER
FOR INFORMATION AND MEDIA

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)

Web searchers encounter following problems

- Old problem - vocabulary mismatch
  - – hard to guess the best search keys; leads to loss of search effectiveness
  - – especially in foreign languages
  - – hard to know word forms (tokens / lemmas)
- Other problems – depending on one's search environment
  - – collection dependency, metadata dependency
  - – engine and query language dependency

TRIM
TAMPERE RESEARCH CENTER FOR INFORMATION AND MEDIA

Background (3)

to annotate

or

not to annotate

Design of WebExplorer

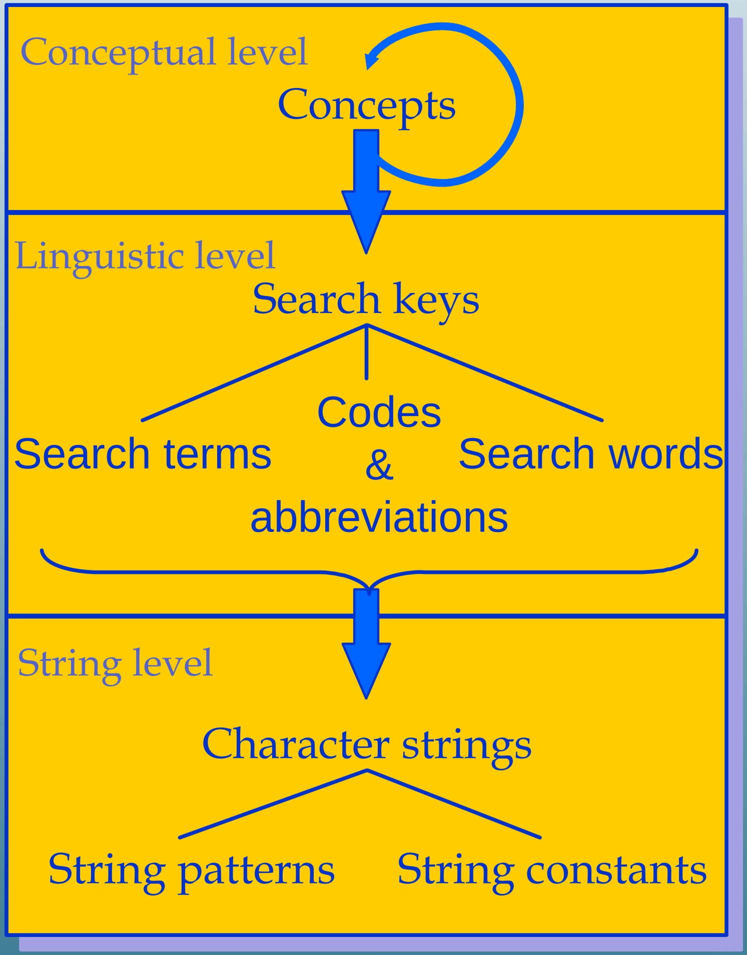
Requirements on the Ontology

- **Personal**
- **Small-scale**
- **Mapping**
- **Multilingual**
- **Editable**

TRIM
TAMPERE RESEARCH CENTER FOR INFORMATION AND MEDIA

Ontologies of WebExplorer based on

Three-level architecture



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

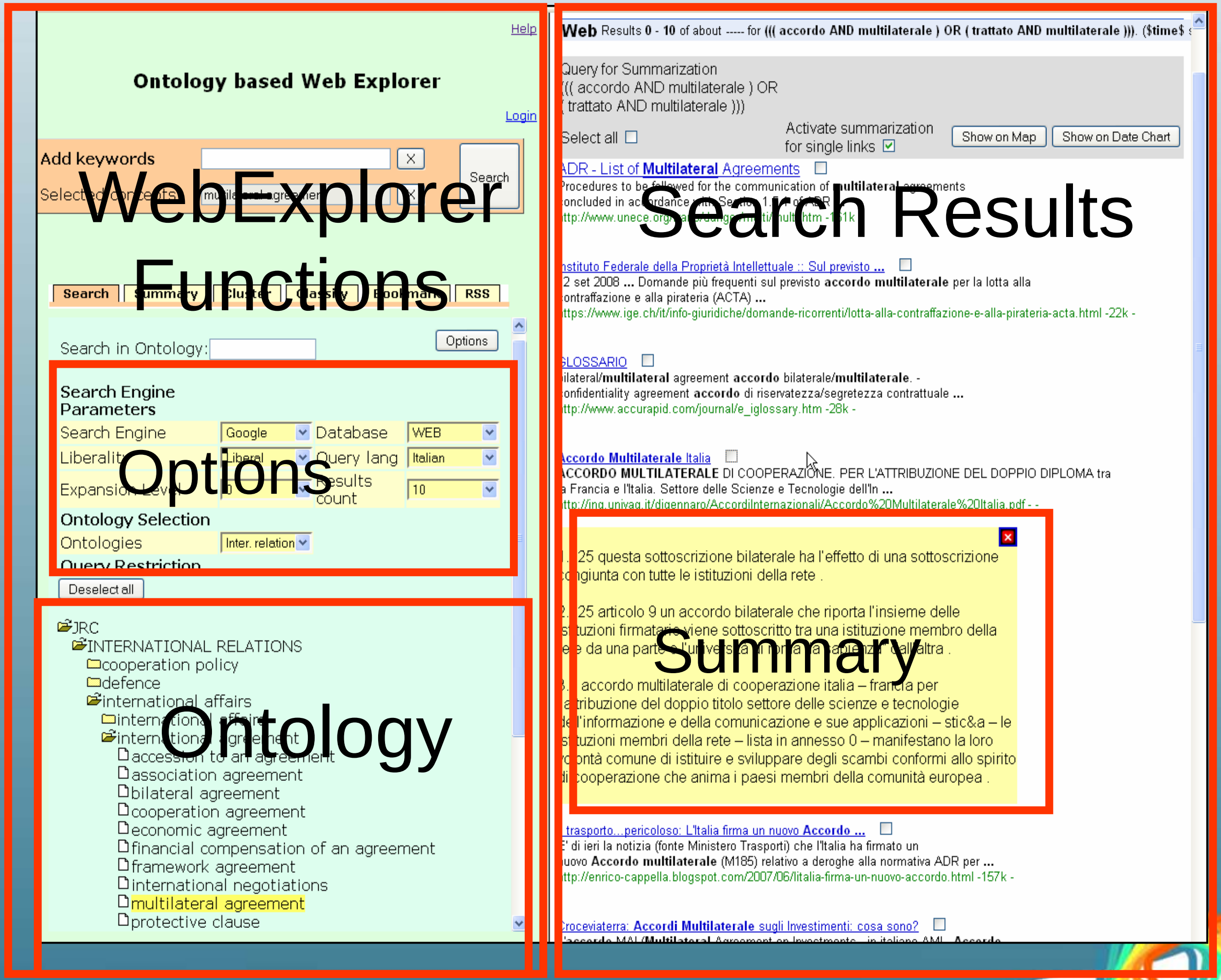
TRIM
TAMPERE RESEARCH CENTER FOR INFORMATION AND MEDIA

Design of WebExplorer

Additional functionalities are required for information exploration:

- **Keyword search**
- **Query-biased summaries**
- **Classification**
- **Clustering**
- **Bookmarking**
- **Smooth integration of functions**

WebExplorer Interface



WebExplorer Functions

Options

Ontology

Search Results

Summary

Help

Ontology based Web Explorer

Login

Add keywords ✕ Search

Selected concepts: multilateral agreement ✕

Search | Summary | Cluster | Classify | Bookmark | RSS

Search in Ontology: Options

Search Engine Parameters

Search Engine	Google	Database	WEB
Liberality	Liberal	Query lang	Italian
Expansion Level	0	Results count	10

Ontology Selection

Ontologies: Inter. relation

Query Restriction

Deselect all

- JRC
  - INTERNATIONAL RELATIONS
    - cooperation policy
    - defence
    - international affairs
      - international affairs
      - international agreement
        - accession to an agreement
        - association agreement
        - bilateral agreement
        - cooperation agreement
        - economic agreement
        - financial compensation of an agreement
        - framework agreement
        - international negotiations
        - multilateral agreement
        - protective clause

Web Results **0** - **10** of about ----- for **(((accordo AND multilaterale) OR (trattato AND multilaterale)))**. ($time$

Query for Summarization
(((accordo AND multilaterale) OR
(trattato AND multilaterale)))

Select all □ Activate summarization for single links ☑ Show on Map Show on Date Chart

ADR - List of **Multilateral** Agreements □
Procedures to be followed for the communication of **multilateral** agreements concluded in accordance with Section 1.5.1 of ADR ...
http://www.unece.org/trans/danger/multi/multi.htm -151k

Istituto Federale della Proprietà Intellettuale :: Sul previsto **...** □
12 set 2008 **...** Domande più frequenti sul previsto **accordo multilaterale** per la lotta alla contraffazione e alla pirateria (ACTA) **...**
https://www.ige.ch/it/info-giuridiche/domande-ricorrenti/lotta-alla-contraffazione-e-alla-pirateria-acta.html -22k -

GLOSSARIO □
bilateral/**multilateral** agreement **accordo** bilaterale/**multilaterale**. -
confidentiality agreement **accordo** di riservatezza/segretezza contrattuale **...**
http://www.accurapid.com/journal/e_iglossary.htm -28k -

Accordo Multilaterale Italia □
ACCORDO MULTILATERALE DI COOPERAZIONE. PER L'ATTRIBUZIONE DEL DOPPIO DIPLOMA tra la Francia e l'Italia. Settore delle Scienze e Tecnologie dell'In **...**
http://ing.univaq.it/digennaro/AccordiInternazionali/Accordo%20Multilaterale%20Italia.pdf - -

☒

1. 25 questa sottoscrizione bilaterale ha l'effetto di una sottoscrizione congiunta con tutte le istituzioni della rete .

2. 25 articolo 9 un accordo bilaterale che riporta l'insieme delle istituzioni firmatarie viene sottoscritto tra una istituzione membro della rete da una parte e l'università di roma la sapienza dall'altra .

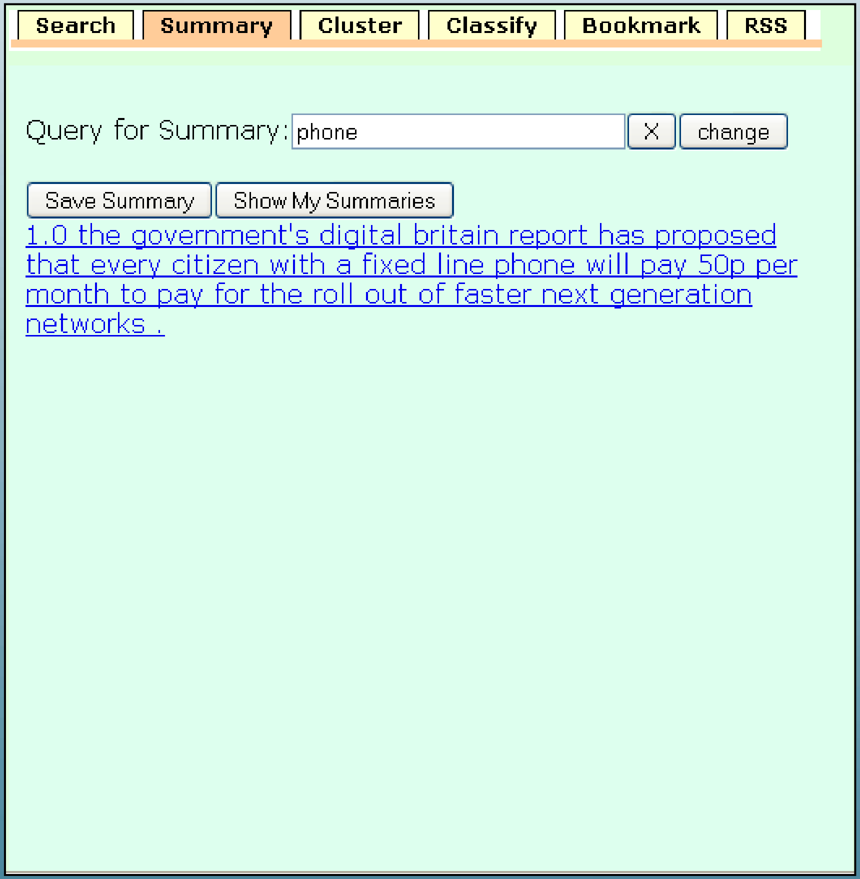
3. accordo multilaterale di cooperazione italia – francia per attribuzione del doppio titolo settore delle scienze e tecnologie dell'informazione e della comunicazione e sue applicazioni – stic&a – le istituzioni membri della rete – lista in annesso 0 – manifestano la loro volontà comune di istituire e sviluppare degli scambi conformi allo spirito di cooperazione che anima i paesi membri della comunità europea .

Il trasporto...pericoloso: L'Italia firma un nuovo **Accordo ...** □
E' di ieri la notizia (fonte Ministero Trasporti) che l'Italia ha firmato un nuovo **Accordo multilaterale** (M185) relativo a deroghe alla normativa ADR per **...**
http://enrico-cappella.blogspot.com/2007/06/litalia-firma-un-nuovo-accordo.html -157k -

Croceviaterra: **Accordi Multilaterale** sugli Investimenti: cosa sono? □

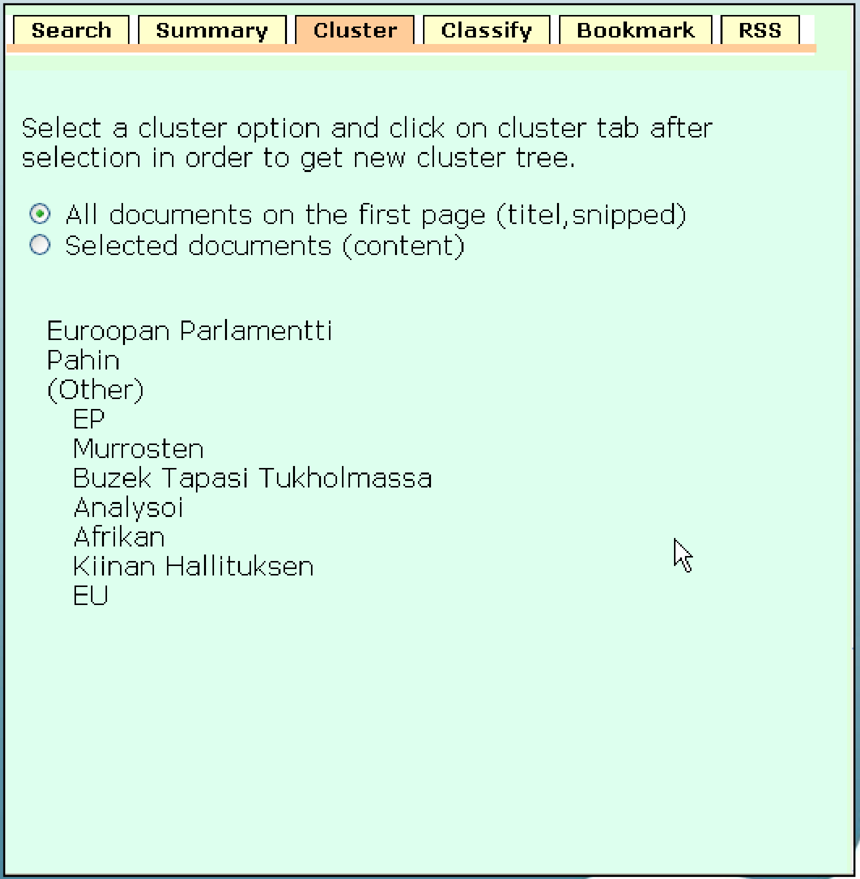
The WebExplorer user interface

WebExplorer Interface (2)



Query for Summary: phone

1.0 the government's digital britain report has proposed that every citizen with a fixed line phone will pay 50p per month to pay for the roll out of faster next generation networks .

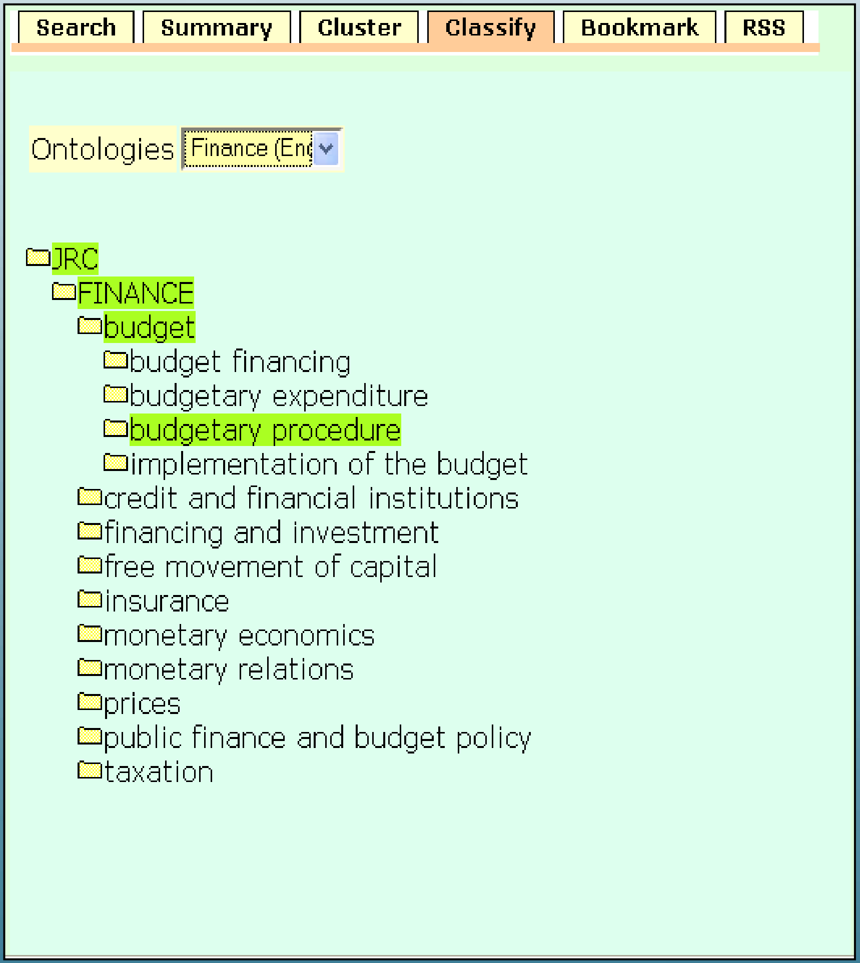


Select a cluster option and click on cluster tab after selection in order to get new cluster tree.

- ● All documents on the first page (titel,snipped)
- ○ Selected documents (content)

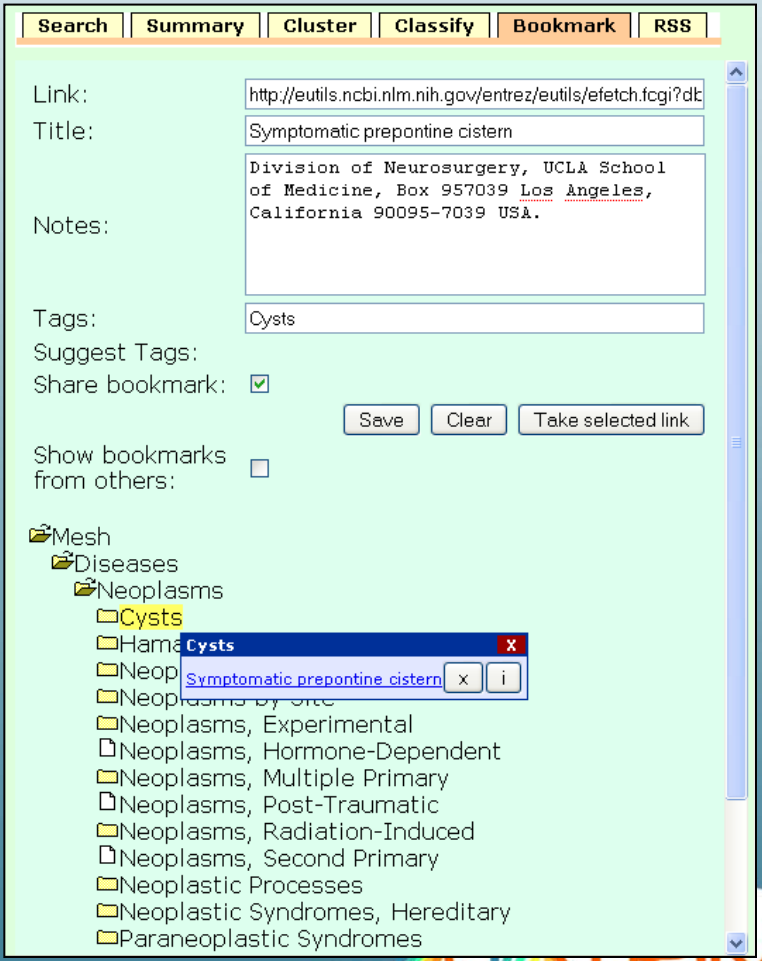
Euroopan Parlamentti
Pahin
(Other)
  EP
  Murrosten
  Buzek Tapasi Tukholmassa
  Analysoi
  Afrikan
  Kiinan Hallituksen
  EU

WebExplorer Interface (3)



Ontologies Finance (Eng

- JRC
  - FINANCE
    - budget
      - budget financing
      - budgetary expenditure
      - budgetary procedure
      - implementation of the budget
    - credit and financial institutions
    - financing and investment
    - free movement of capital
    - insurance
    - monetary economics
    - monetary relations
    - prices
    - public finance and budget policy
    - taxation



Link: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db

Title: Symptomatic prepontine cistern

Notes: Division of Neurosurgery, UCLA School of Medicine, Box 957039 Los Angeles, California 90095-7039 USA.

Tags: Cysts

Suggest Tags:

Share bookmark: ☑

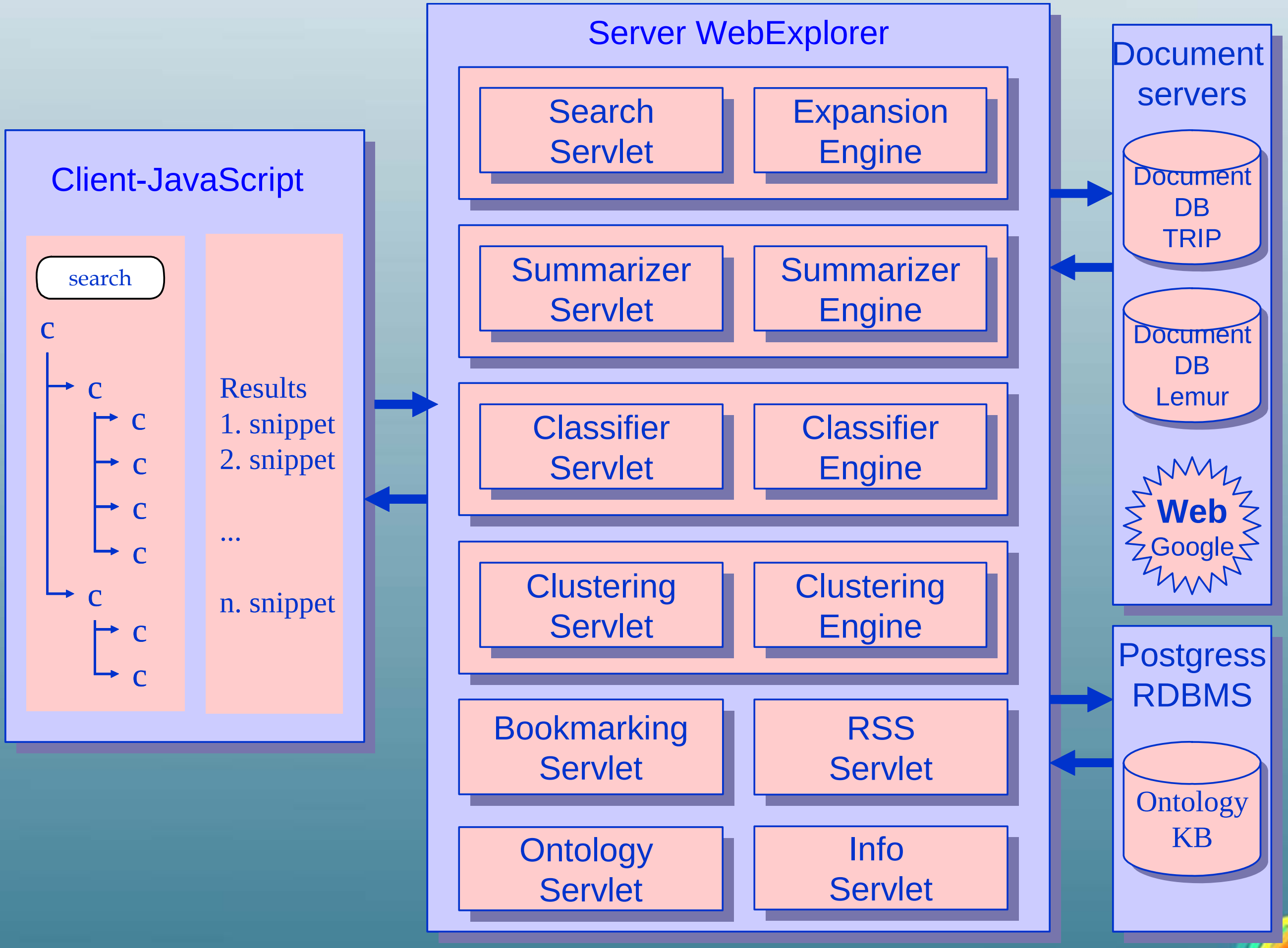
Show bookmarks from others: □

- Mesh
  - Diseases
    - Neoplasms
      - Cysts
      - Hama
      - Neop
      - Neoplasms by Site
      - Neoplasms, Experimental
      - Neoplasms, Hormone-Dependent
      - Neoplasms, Multiple Primary
      - Neoplasms, Post-Traumatic
      - Neoplasms, Radiation-Induced
      - Neoplasms, Second Primary
      - Neoplastic Processes
      - Neoplastic Syndromes, Hereditary
      - Paraneoplastic Syndromes

Cysts X

Symptomatic prepontine cistern x i

WebExplorer System Architecture



WebExplorer System Architecture Diagram

Formative Evaluation

To test the feasibility of WebExplorer

	Group 1	Group 2
Ontology IR	+2	+1
CL-IR	+2	+1.5
Summarization	0	+1
Classification	+0.5	--
Bookmarking	+1.5	0.5
RSS Feeds	+0.5	--

Median values of the groups' opinions

Scale: -2 to +2

-2:	No benefit at all
-1:	Only little benefit
0:	I can't say
+1:	some benefit
+2:	much benefit

School grade (4-10) for WebExplorer:7

Conclusion

- WebExplorer is an answer to problems in semantic information access
  - **light-weight** disposable **ontologies** for direct 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**
  - annotation - media analysis
  - a compromise different from semantic annotation or library indexing with control at the user end
    - costs and benefits distributed in a new way

Conclusion

Document Analysis functions offer

- Online query biased **summarization** of individual documents and document sets
- **Clustering** of documents
- Ontology based cross-lingual **classification**
- **Bookmark** managemet
- Displaying search results on **map** and **timeline**

Thank you!

Any Questions?

TRIM
TAMPERE RESEARCH CENTER FOR INFORMATION AND MEDIA