OmniPaper
The smartest European news finder

Elpub 2004: workshop 4

Advanced search techniques based on an EuroWordNet application, through demonstrations of the OmniPaper project results, and technological approaches as implemented in the OmniPaper prototype, including SOAP and WSDL

Jan Engelen, Katholieke Universiteit Leuven, Belgium
Bert Paepen, Katholieke Universiteit Leuven, Belgium
Markus Schranz, Pressetext, Austria

Contents

→ OmniPaper project introduction
- "Local Knowledge Layer" prototypes:
  - Automatic Keyword Extraction
  - SOAP
  - RDF
  - Topic Maps
- Comparison of prototype performance
- "Overall Knowledge Layer" architecture
- "Overall Knowledge Layer" prototype
- Conclusions and future work

Project objectives

- Starting points:
  - Too much information
  - Coupling remains difficult
- Propose ways for overcoming these problems
- Make information:
  - Easier and faster accessible
  - More accurate
  - Increasingly valuable using self-learning

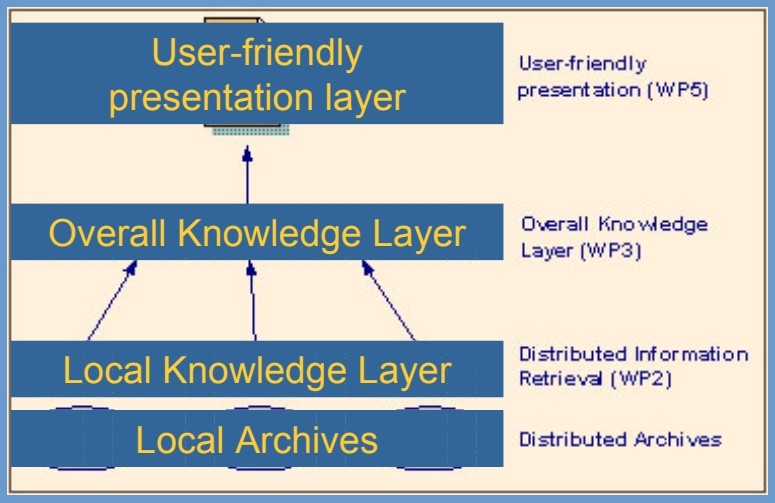
Project objectives

- Combine shattered information sources so that the total is greater than the sum
  "virtual super-archive"
- Key objective:
  "Creation of a multilingual navigation and linking layer on top of distributed information resources in a self-learning environment"

Problem areas and solutions

- Information exchange problem: XML
  - SOAP: universal interface to distributed archives
  - RDF: model for expressing meta-data
  - Topic Maps: navigation, indexing, linking
- Self-learning behaviour
  - Keyword extraction: *data mining*

Project architecture



Project architecture

- Local archives
- Local Knowledge Layer
  - Smart search mechanisms on one archive in one language
  - How to access local archive (SOAP)
  - How to enrich information with keywords (AKE)
  - How to handle metadata (RDF)
  - How to link keywords and metadata for searching (Topic Maps)
- Overall Knowledge Layer
  - Multi-archive
  - Multi-lingual
- User-friendly presentation layer

Blueprint

- Provide possible solutions at a **general level**
  - Any electronic archive that contains cultural, scientific information
    - document management systems
    - digital museums
    - medical databases
  - Constraint: text oriented
- Blueprint must provide general answers
  - Cross-evaluation

Newspaper prototype

- Proof of concept for general solutions offered by blueprint
- Work at technical, content and language issues at a more detailed level
- Good example of distributed information in many languages, computer systems, databases, ...

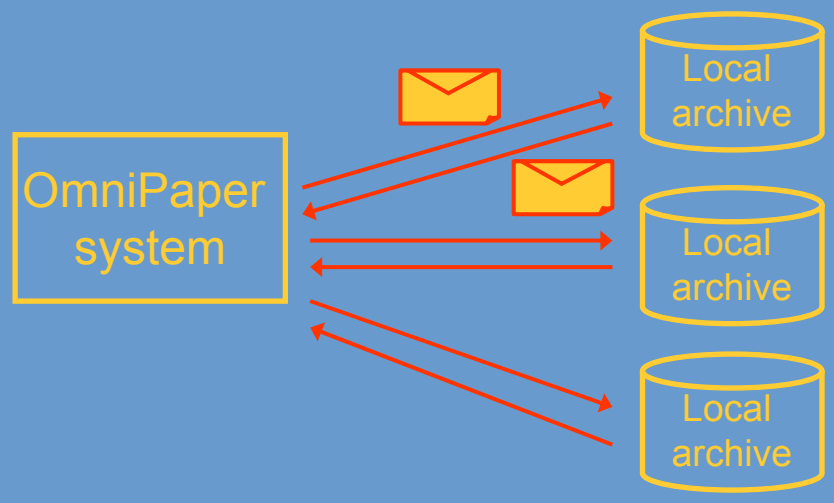
Contents

- OmniPaper project introduction
- → "Local Knowledge Layer" prototypes:
  - → SOAP
  - Automatic Keyword Extraction
  - RDF
  - Topic Maps
- Comparison of prototype performance
- "Overall Knowledge Layer" architecture
- "Overall Knowledge Layer" prototype
- Conclusions and future work

SOAP

- Simple Object Access Protocol
  - Protocol for information exchange across the Internet
  - Uses HTTP
    - Passes through firewalls
  - Uses XML as message format
- Application-, platform- and hardware-independent
- Messages are packed in “envelopes”

SOAP and OmniPaper



WSDL

- Web Services Definition Language
  - XML language for describing web services
  - Describes:
    - Message vocabulary (XML Schema)
    - Messages and their order
    - Binding to SOAP or other protocols

WSDL

- OmniPaper SOAP service operations:
  - FullTextSearch
    - Full-text search in entire contents of article
  - ParameterSearch
    - Full-text search in given metadata fields
  - IdentifiedSearch
    - Retrieves one news article using its ID
  - NewsUpdate
    - Notifies OmniPaper system about newly added articles (! From local archive to OmniPaper)
  - MetaDataFeedback
    - Retrieves metadata of given article (! From local archive to OmniPaper)
    - Goal: update local archives with OmniPaper info
  - MetaDataReady
    - Notifies the local archive that the metadata of newly added or updated news articles is ready
  - NewsSolicit
    - Asks the local archive if new (or updated) articles are available

SOAP prototype

- SOAP server: My News archive in Barcelona
- SOAP client: prototype in Leuven
- Client offers an alternative search interface to the My News interface
- Search functionality provided by My News back-end

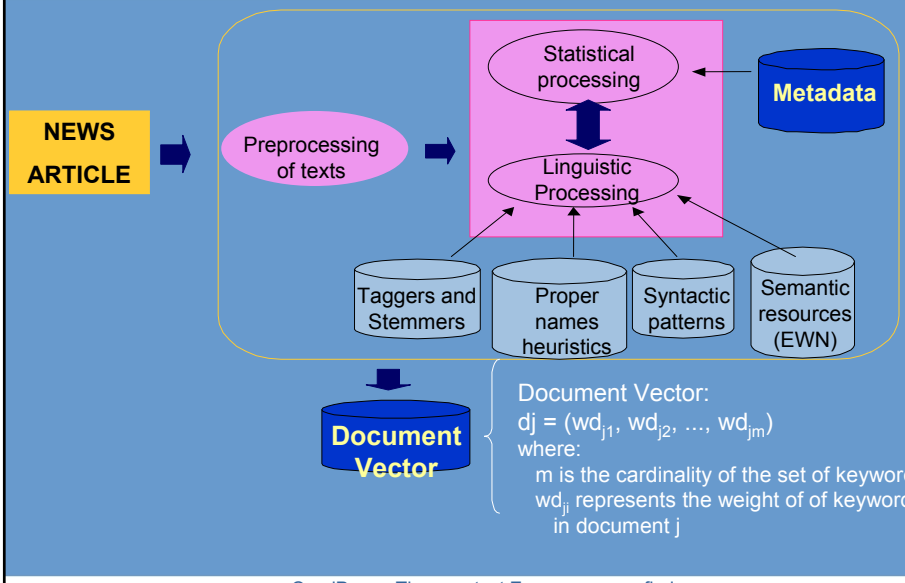
Contents

- OmniPaper project introduction
- → "Local Knowledge Layer" prototypes:
  - SOAP
  - → Automatic Keyword Extraction (AKE)
  - RDF
  - Topic Maps
- Comparison of prototype performance
- "Overall Knowledge Layer" architecture
- "Overall Knowledge Layer" prototype
- Conclusions and future work

Automatic Keyword Extraction

- Subsystem for automatically extracting keywords from news articles
- Linguistical and statistical methods to:
  - Eliminate stop words
  - Identify "keywords": meaningful words with regard to the article content
  - Apply weights to keywords
    - Vector of weighted keywords per article
- Extracted keywords are used in the RDF and Topic Map prototypes
- Keywords can be manually corrected using the "AKE workbench"

AKE process



Document Vector:

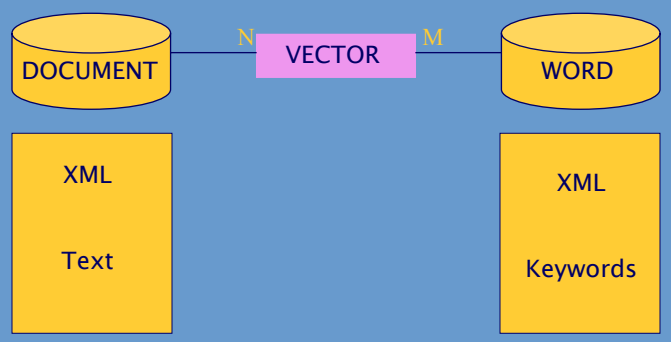
$dj = (wd_{j1}, wd_{j2}, ..., wd_{jm})$

where:

m is the cardinality of the set of keywords

wd_{ji} represents the weight of of keyword i in document j

AKE Prototype



Demo (I): Input Document

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="yes" ?>
<documento>
<articulo id="DTG200209060160">
  <head>
    <DREDATE>06/09/2002</DREDATE>
    <Newspaper>THE DAILY TELEGRAPH</Newspaper>
    <Title>We &apos;re forgotten,  say victims of  McVeigh bomb
   America One Year On   </Title>
    <Author> By Oliver Poole in Oklahoma City</Author>
    <Page>16</Page>
    <Perfil>The_Daily_Telegraph</Perfil>
  </head>
  <body>
    <DRECONTENT><p>
   IN A small house in a rundown area of Oklahoma City, with
   welfare cheques lying on the kitchen table, a young boy sits in
   the corner, speaking through a hole cut in his throat to allow
   him to breathe.</p>
    <p> Such is the daily reality of life for a victim of the
   &apos;wrong&apos; terrorist attack.</p>
```

Demo (I): Input Document


```
   <p> Seven years ago P J Allen was a two-year-old boy at a day
care centre on the second floor of the Alfred P Murrah Building
when the federal offices were blown up by home-grown terrorist
hoping to launch a race war...</p>
    <p> P J &apos;s lungs were scorched by the heat from the bomb
Timothy McVeigh detonated. He was given a permanent tracheotomy
to keep him alive. Doctors have told his parents that he is too
prone to infection ever to attend school.</p>
    <p> But unlike those who lost loved ones or suffered horrific
injuries in the terrorist attacks on the World Trade Centre and
the Pentagon a year ago there is no multi-million dollar
compensation pay-out from the government for him.</p>
    <p> Instead he and his grandmother, who looks after him,
survive on a monthly pounds 340 social security cheque and
charity handouts.</p>
    <p> B</p>
    <p>  {PS}News:  {ES}  International:  War on terrorism</p>
   </DRECONTENT>
  </body>
</articulo>
</documento>
```


OmniPaper: The smartest European news finder

Demo (I): Resulting Keywords


```
  <key-list>
  <keyword weight="18.593901" df="150" tf="4">attack</keyword>
  <keyword weight="17.050301" df="196"
  tf="4">September</keyword>
  <keyword weight="15.833300" df="604" tf="6">people</keyword>
  <keyword weight="13.837700" df="342"
  tf="4">government</keyword>
  <keyword weight="12.899600" df="191" tf="3">Centre</keyword>
  <keyword weight="10.180400" df="358" tf="3">ago</keyword>
  <keyword weight="9.964930" df="119" tf="2">York</keyword>
  <keyword weight="9.916840" df="121"
  tf="2">received</keyword>
  <keyword weight="9.687720" df="131"
  tf="2">building</keyword>
  <keyword weight="9.622390" df="134" tf="2">special</keyword>
  ...
  <keyword weight="2.905740" df="502" tf="1">good</keyword>
  <keyword weight="2.849380" df="522" tf="1">world</keyword>
  <keyword weight="2.638880" df="604" tf="1">made</keyword>
  <keyword weight="2.182060" df="829" tf="1">time</keyword>
```


OmniPaper: The smartest European news finder

Demo (II): Input Document

```
?xml version="1.0" encoding="ISO-8859-1" standalone="yes" ?>
<documento>
<articulo id="DTG200209300145">
  <head>
    <DREDATE>30/09/2002</DREDATE>
    <Newspaper>THE DAILY TELEGRAPH</Newspaper>
    <Section>DEPORTES Sport , Soccer ,</Section>
    <Title>Leicester show grit   </Title>
    <Author> By Ben Findon</Author>
    <Page>06</Page>
    <Perfil>The_Daily_Telegraph</Perfil>
  </head>
  <body>
   <DRECONTENT>
     <p> LEICESTER CITY 1  WOLVERHAMPTON 0   FROM
Leicester in the east to Wolverhampton in the west, the Midlands
 seems a division within a division. All those frantic derby
 battles complicating the wider promotion picture.</p>
     <p> It was trench warfare at the Walkers Stadium, where
 resolve flowed from every pore in industrial-strength
 quantities.</p>
      <p>Yellow cards - and one red - peppered the afternoon and
 statistics showed Leicester winning without managing a shot on
 target.</p>
```

OmniPaper: The smartest European news finder

Demo (II): Input Document

```
     <p> That is discounting the penalty gained and dispatched by
Paul Dickov in the sixth minute, but Wolves should not feel too
indignant: they could muster only four goalbound efforts.</p>
     <p> &apos;Not one of our more fluent performances,&apos;
conceded Leicester manager Micky Adams, perhaps acknowledging...
     ...
     <p> Leicester City (4-4-2): Walker; Sinclair, Elliott,
Taggart, Davidson; Scowcroft, McKinlay, Izzet, Rogers (Summerbee
75); Dickov (Benjamin 83), Deane (Reeves 69). Subs: Heath,
Ashton. Goal: Dickov (6 pen). Booked: Izzet, Sinclair, Taggart,
Dickov. Sent off: Izzet.</p>
     <p> Wolverhampton (4-4-2): Murray; Irwin, Butler, Lescott,
Naylor; Newton, Rae, Ince, Cooper (Ndah 72); Sturridge (Miller
77), Blake. Subs: Solly, Ingimarsson, Edworthy.</p>
     <p>Booked: Rae, Ince.</p>
     <p> Referee: M Riley (Leeds).</p>
    </DRECONTENT>
  </body>
</articulo>
</documento>
```

OmniPaper: The smartest European news finder

Demo (II): Results Compared

- applying vs. not applying stemming:

Without Stemming

```
<key-list>
...
<keyword weight="5.24"> leaving </keyword>
<keyword weight="5.00"> showed </keyword>
<keyword weight="5.00"> body </keyword>
...
<keyword weight="4.90"> received </keyword>
...
<keyword weight="4.80"> minute </keyword>
<keyword weight="4.45"> minutes </keyword>
....
</key-list>
```

With Stemming

```
<key-list>
...
<keyword weight="3.80"> leav </keyword>
<keyword weight="6.00"> show </keyword>
<keyword weight="4.58"> bodi </keyword>
...
<keyword weight="4.00"> receiv </keyword>
...
<keyword weight="7.61"> minut </keyword>
...
...
</key-list>
```


Contents

- OmniPaper project introduction
- → "Local Knowledge Layer" prototypes:
  - SOAP
  - Automatic Keyword Extraction (AKE)
  - → RDF
  - Topic Maps
- Comparison of prototype performance
- "Overall Knowledge Layer" architecture
- "Overall Knowledge Layer" prototype
- Conclusions and future work

RDF

- Resource Description Framework
- RDFMSS – W3C recommendation since February 1999
- RDFSS – W3C candidate recommendation since March 2000

RDF

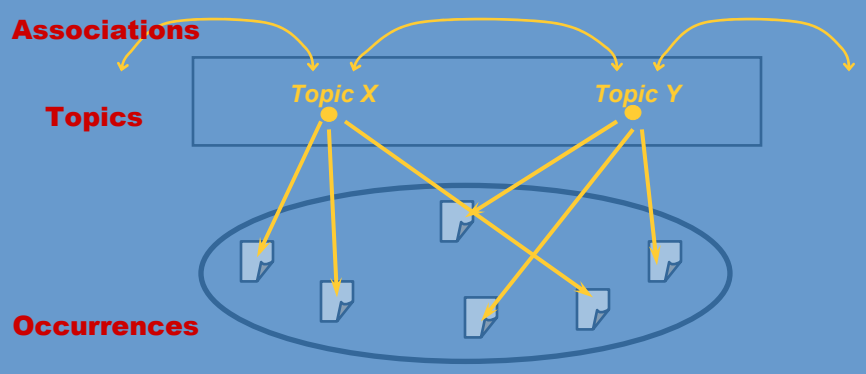
- RDF = used in OmniPaper for encoding metadata
- Metadatabase: native RDF database (RDF Gateway)
- RDF prototype:
  - Searches for news articles using metadata fields
  - Simple search (not using metadata fields) is performed on extracted keywords
    - "pure keyword search"
    - Uses full-text search on keywords per article
  - Results are ranked by relevance

Contents

- OmniPaper project introduction
- → "Local Knowledge Layer" prototypes:
  - SOAP
  - Automatic Keyword Extraction (AKE)
  - RDF
  - → Topic Maps
- Comparison of prototype performance
- "Overall Knowledge Layer" architecture
- "Overall Knowledge Layer" prototype
- Conclusions and future work

Topic Maps = Linking + Indexing

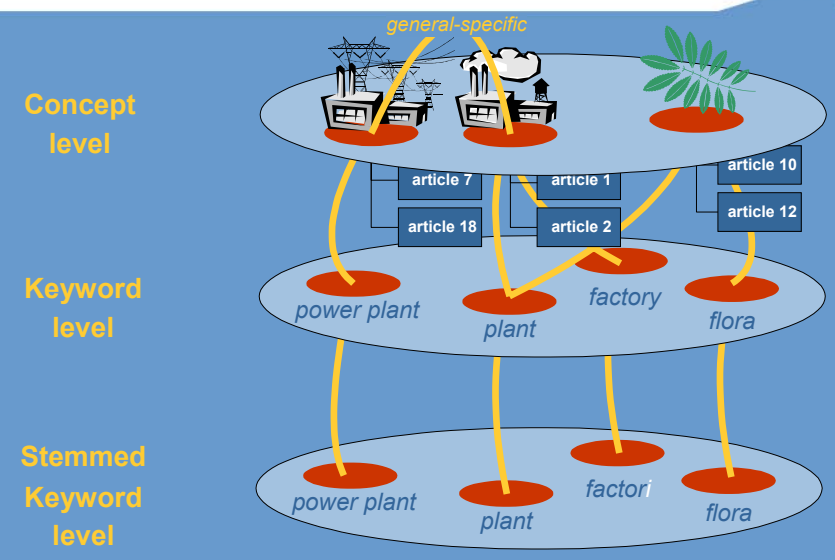
- Topic Map Model: XML syntax (XTM 1.0 DTD)
- The TAO of Topic Maps: Topics, Occurrences, Associations



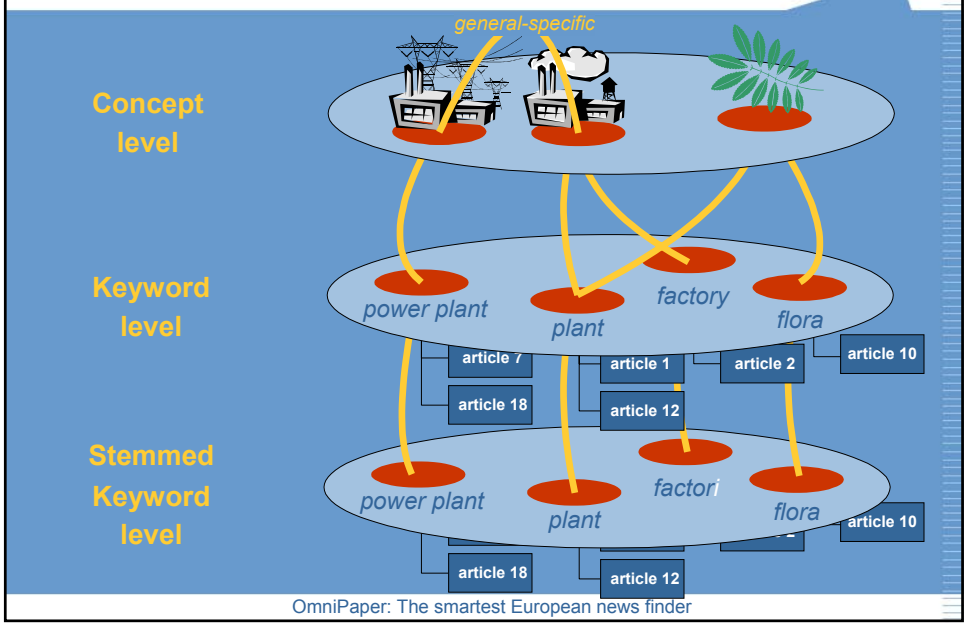
Topic Maps and OmniPaper

- Topics
  - Digital representation of real-word concepts
  - Keywords and stemmed keywords
  - Proper names, …
- Associations
  - (Semantic) relations between topics
- Occurrences
  - Articles

Topic Maps and OmniPaper



Topic Maps and OmniPaper



OmniPaper: The smartest European news finder

OmniPaper and WordNet

- WordNet = linguistic resource for these concepts, keywords and relations (English only)
- Concept = “synset” from WordNet
- Keyword = words describing a synset
  - Most important keyword for a concept is called *preferred keyword*
- Semantic relations between concepts
  - Dozens of relation types
  - Summarized into:
    - Equivalence relation
    - General-specific relation (two-way)
    - Association

OmniPaper: The smartest European news finder

From query to results

1. Enter a new query

`(biochemical OR weapon OR nuclear) AND Iraq`

- Possibly with advanced search options

OmniPaper: The smartest European news finder

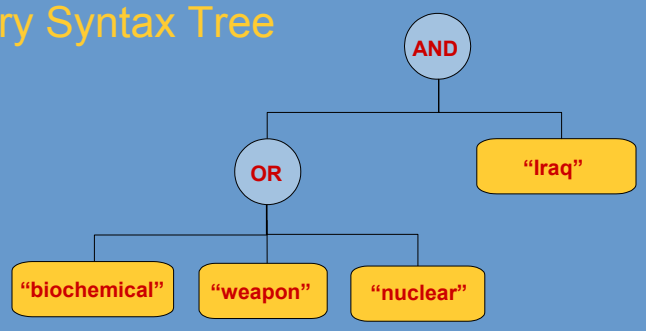
From query to results

1. Enter a new query

`(biochemical OR weapon OR nuclear) AND Iraq`

- Possibly with advanced search options

2. Query Syntax Tree



OmniPaper: The smartest European news finder

From query to results

1. Enter a new query

`(biochemical OR weapon OR nuclear) AND Iraq`

   - Possibly with advanced search options
2. Query Syntax Tree
3. Find the concepts in the topic map that correspond to the basic search terms

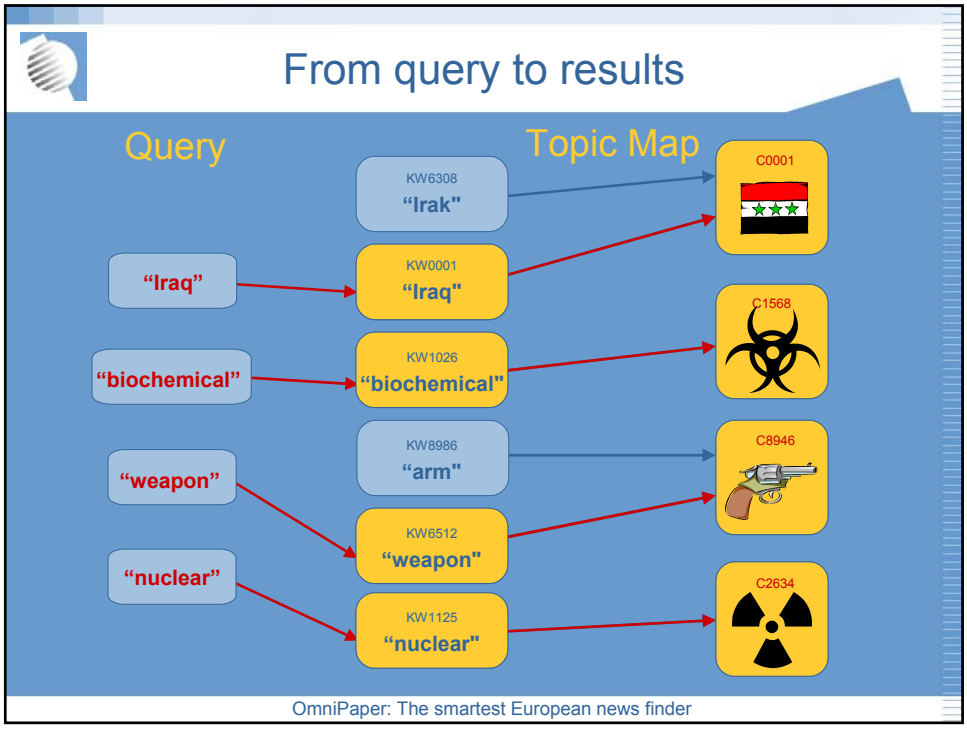
From query to results

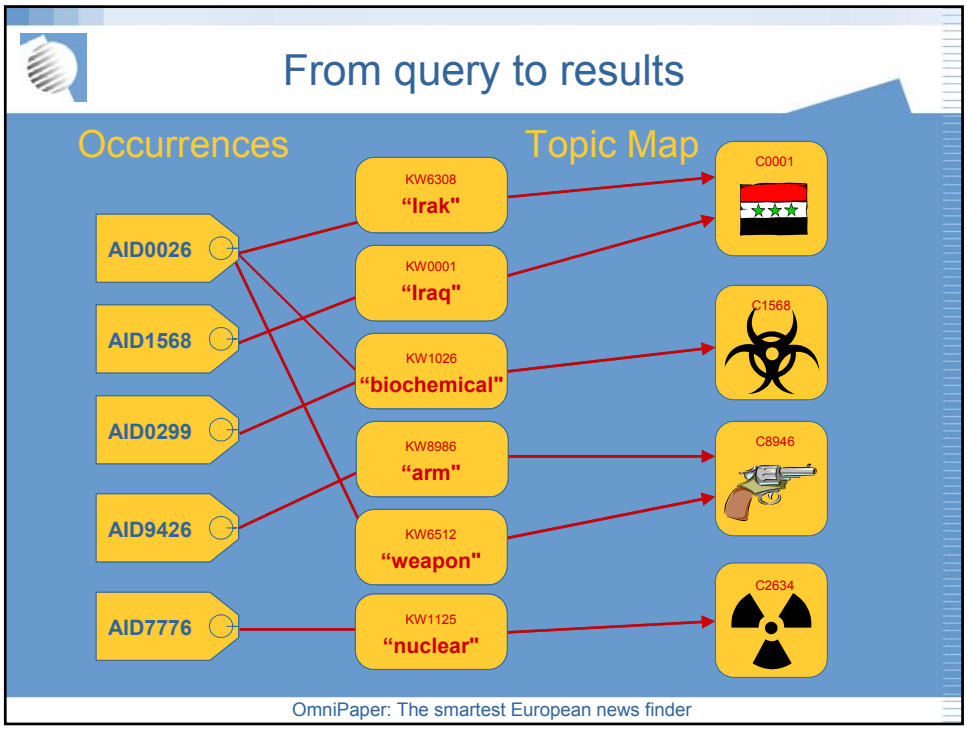
1. Enter a new query

`(biochemical OR weapon OR nuclear) AND Iraq`

   - Possibly with advanced search options
2. Query Syntax Tree
3. Find the concepts in the topic map that correspond to the basic search terms
4. Find basic results for each basic query

OmniPaper: The smartest European news finder


From query to results
Query
Topic Map
"Iraq"
"biochemical"
"weapon"
"nuclear"
KW6308
"Irak"
KW0001
"Iraq"
KW1026
"biochemical"
KW8986
"arm"
KW6512
"weapon"
KW1125
"nuclear"
C0001
C1568
C8946
C2634
OmniPaper: The smartest European news finder


From query to results
Occurrences
Topic Map
AID0026
AID1568
AID0299
AID9426
AID7776
KW6308
"Irak"
KW0001
"Iraq"
KW1026
"biochemical"
KW8986
"arm"
KW6512
"weapon"
KW1125
"nuclear"
C0001
C1568
C8946
C2634
OmniPaper: The smartest European news finder

From query to results

1. Enter a new query

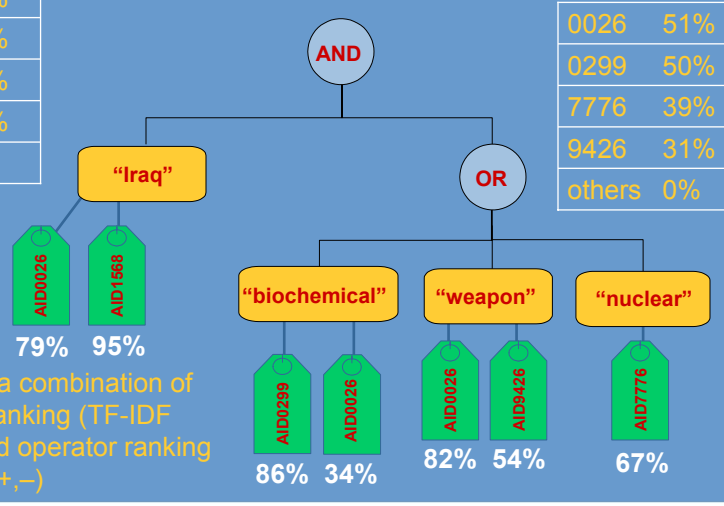
```
(biochemical OR weapon OR nuclear) AND Iraq
```

   - Possibly with advanced search options
2. Query Syntax Tree
3. Find the concepts in the topic map that correspond to the basic search terms
4. Find basic results for each basic query
5. Recombine basic results according to Query Syntax Tree

From query to results

0026	62%
1568	29%
0299	21%
7776	17%
9426	14%
others	0%

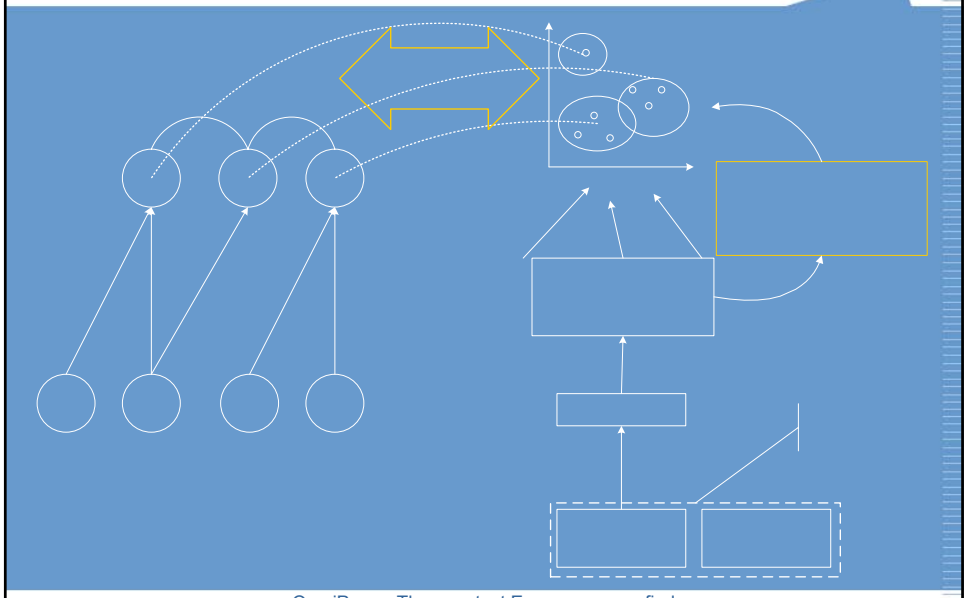
Extended Boolan Model



0026	51%
0299	50%
7776	39%
9426	31%
others	0%

Based on a combination of keyword ranking (TF-IDF model) and operator ranking (AND,OR,+,–)

Relation Topic Maps - AKE



Contents

- OmniPaper project introduction
- "Local Knowledge Layer" prototypes:
  - SOAP
  - Automatic Keyword Extraction (AKE)
  - RDF
  - Topic Maps
- → Comparison of prototype performance
- "Overall Knowledge Layer" architecture
- "Overall Knowledge Layer" prototype
- Conclusions and future work

Comparison of prototype performance

- Main comparison criteria:
  - Relevancy:
    - Recall: ability to return all relevant documents
    - Precision: ability to return only relevant documents
    - F-value (sort of summary between recall & precision)
  - Response time
- Data set:
  - 1881 English news articles (Sep 2002)
- Test sets
  - Manually created:
    - 18 topics
    - 45 queries
    - corresponding reference result sets
- Test tool: Automatic Test Engine

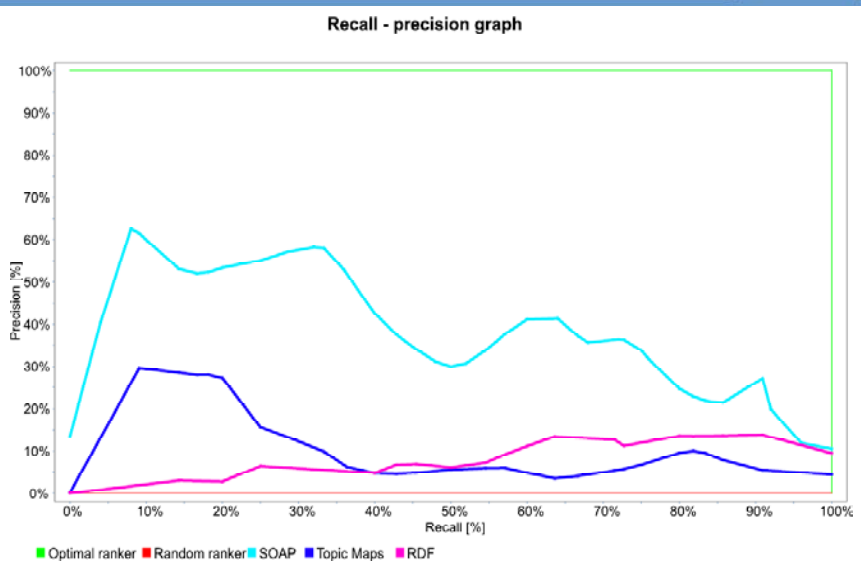
Test results: relevancy

Relevancy summary table:

	SOAP	RDF	Topic Maps
Recall	0.525	0.334	0.316
Precision	0.479	0.272	0.071
F-value	0.447	0.193	0.105

(values are averages over queries)

Test results: relevancy



Test results: relevancy

- Low average precision
  - Prototypes return too many bad results
  - High precision for low recall?
    - SOAP: best
    - Topic Maps: medium
    - RDF: worse
- Average recall:
    - SOAP: best
    - RDF: medium
    - Topic Maps: worse
    - Reasons:
      - Influence of AKE
      - WordNet: Fine-grainedness + for Topic Maps: words not appearing in WordNet do not yield any results (↔ fulltext searching)

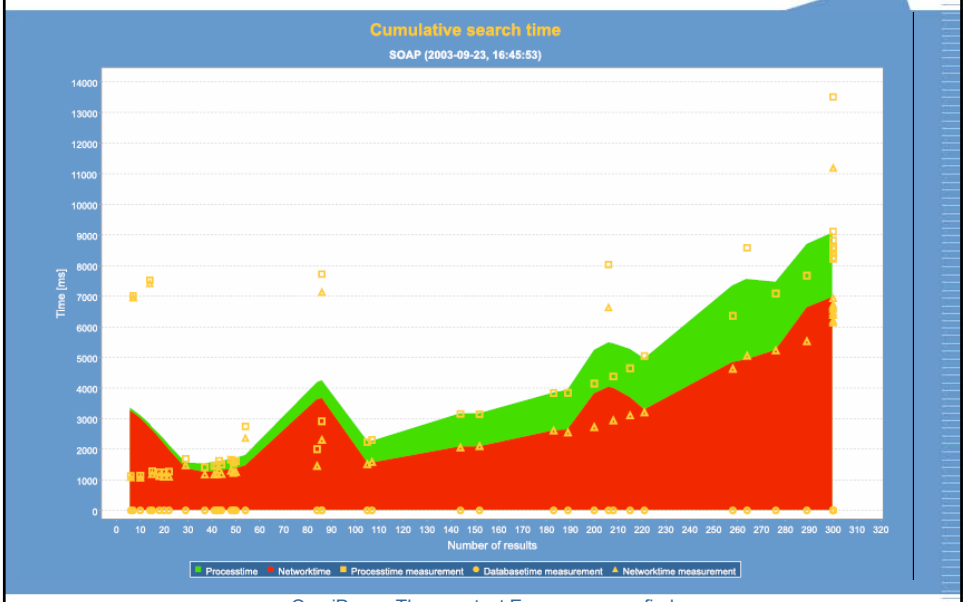
Test results: relevancy

- SOAP vs Topic Maps and RDF:
  - SOAP=better because back-end search system uses Autonomy database (smart indexing)
- Topic Maps precision
  - Average: low
  - Higher at low recall rates → better ranking because of semantic query expansion
- Influence of stemming:
  - Slight gain in recall
  - No gain in precision

Test results: Search time

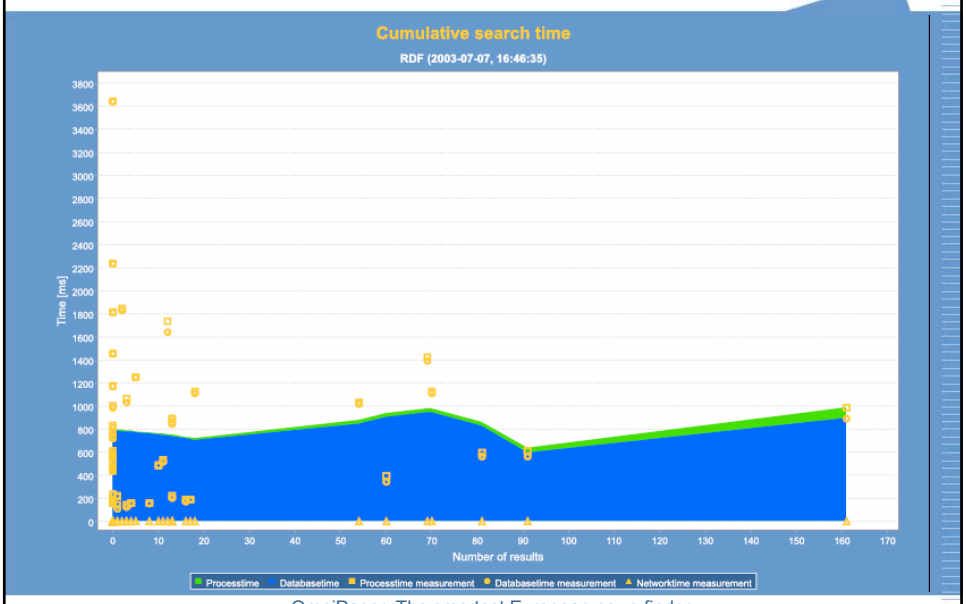
- Average search time per query:
  - RDF: 0.791 s
    - 90% database time
    - Very performant RDF database – in-memory storage
  - SOAP: 4.551 s
    - 80% network time
  - Topic Maps: 4.735 s
    - 95% database time because of WordNet lookups

Search time: SOAP



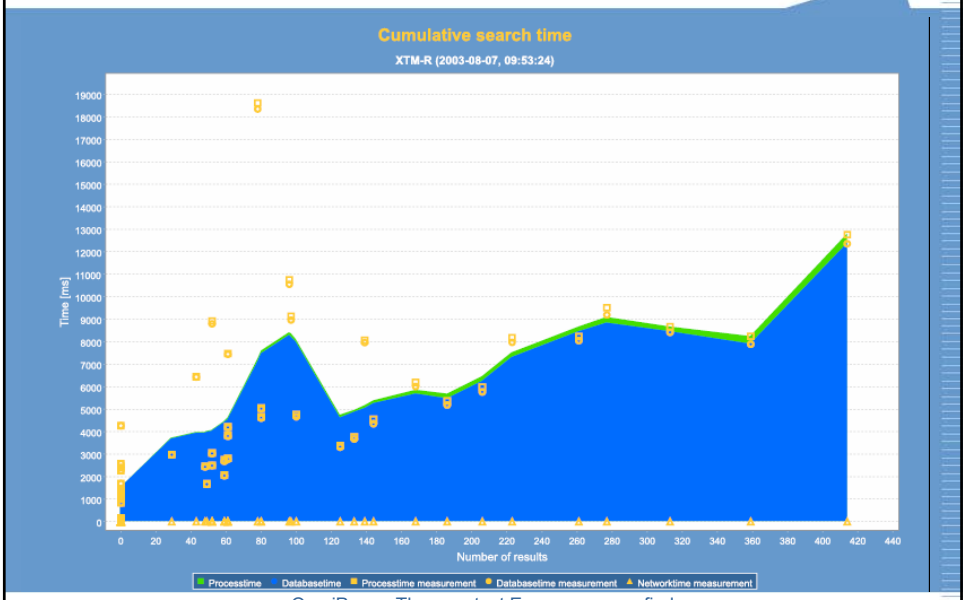
OmniPaper: The smartest European news finder

Search time: RDF



OmniPaper: The smartest European news finder

Search time: Topic Maps



OmniPaper: The smartest European news finder

User evaluation results

- Users are not confident enough in search results
- Users like the query refinement tool (web of concepts)
- Users are not using the hierarchical subject view for searching news articles

OmniPaper: The smartest European news finder

Contents

- OmniPaper project introduction
- "Local Knowledge Layer" prototypes:
  - SOAP
  - Automatic Keyword Extraction (AKE)
  - RDF
  - Topic Maps
- Comparison of prototype performance
- → "Overall Knowledge Layer" architecture
- "Overall Knowledge Layer" prototype
- Conclusions and future work

Contents

- OmniPaper project introduction
- "Local Knowledge Layer" prototypes:
  - SOAP
  - Automatic Keyword Extraction (AKE)
  - RDF
  - Topic Maps
- Comparison of prototype performance
- "Overall Knowledge Layer" architecture
- → "Overall Knowledge Layer" prototype
- Conclusions and future work

OmniPaper: The smartest European news finder

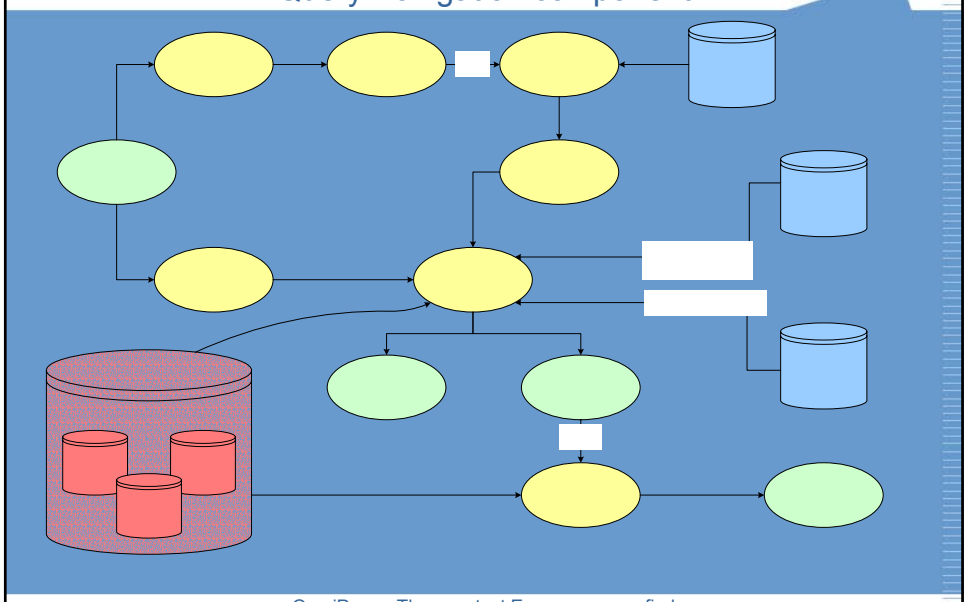
Overall Knowledge Layer prototype

- Local Knowlegde Layer final prototype:
  - Combination of prototypes:
    - SOAP: getting articles from local archive
    - RDF: metadatabase
    - Topic Maps: search and navigation ("web of concepts")
- Overall Knowledge Layer prototype:
  - Combination of SOAP vs Topic Map way of searching
  - Adding multi-archive search
  - Adding multilingual search
  - Dynamic data set: latest news

Overall Knowledge Layer prototype

- Smart search:
  - Not pure Topic Map search
  - Combination of fulltext search (through SOAP) and semantic query expansion (Topic Maps)
- Multilinguality:
  - EuroWordNet database used
    - → Only English, Spanish, German, French and Dutch
  - Query terms translated using ILI (interlingual index)
  - User interface always in one language (except articles)
  - Online translation engine for articles
  - Automatic language identification
- Web of concepts: graphical version
- Multi-archive search

Design
Query/Navigation component



OmniPaper: The smartest European news finder

Design
Query module

- Query engine steps:
  - Query language detection
  - Query keyword extraction (AKE)
  - Query translation
  - Semantic query expansion in all languages (using synonyms)
  - Forwarding of enhanced queries to news brokers

OmniPaper: The smartest European news finder

Qu

User
Request

Navig

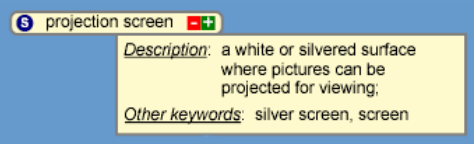
Graphical 'Web of Concepts'

- Functions:
  - Feedback on query:
    - Show information about current terms
    - Semantically related terms
    - Disambiguate query terms *(refine query)*
    - Remove terms from query
  - Explore related terms *(refocus query)*
    - Semantic relations: more general, more specific, equivalent, associated

OmniPaper: The smartest European news finder

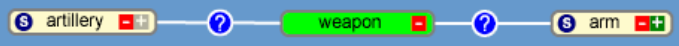
Graphical 'Web of Concepts'

- Items to be represented:
  - Concept



  - Focus Concept

  - Keyword



OmniPaper: The smartest European news finder

Graphical 'Web of Concepts'

- Items to be represented (continued)
  - Non-existing keyword

  - Association

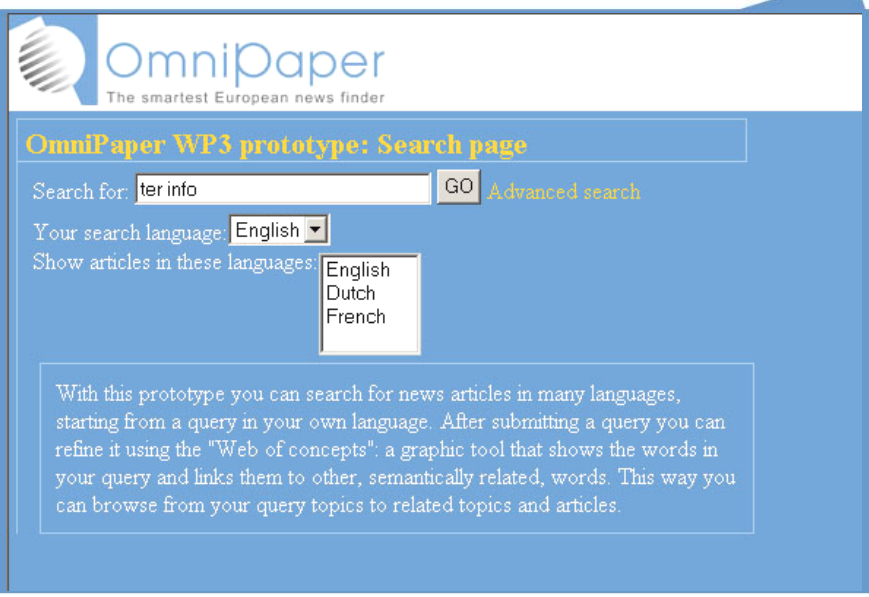
Graphical 'Web of Concepts'

- Features
  - Expand concept
  - Remove topic
  - Select concept
  - History
  - Drag & drop topics
  - Zooming
  - Panning

Graphical 'Web of Concepts'

- Example:
  - Query: "Iraq AND weapon"

Overall Knowledge Layer prototype



Overall Knowledge Layer prototype

OmniPaper
The smartest European news finder

OmniPaper WP3 prototype: Search page

Search for:

Your search language: English

Show articles in these languages: English, Dutch, French

From: January 2002 — To: January 2002

Title:

Subject:

Publisher:

Author:

GO Simple search

OmniPaper: The smartest European news finder

Overall Knowledge Layer prototype

OmniPaper — Search for: dawn — GO — Advanced search

News subjects | Article

Web of concepts

Search results

Your original query: dawn

Show expanded query

40 Results — Results 21 to 25

Hoon scrapped Army equipment because of cost		
0.14%	2004-2-6	THE DAILY TELEGRAPH
DTI rapped for 'ostrich' attitude to nuclear group		
0.14%	2004-2-6	THE DAILY TELEGRAPH
Libreta de polideportivo		
0.14%	2004-2-6	Clarin
More cases in doubt as abuse verdict quashed		
0.14%	2004-2-6	The Guardian
Unos 300 luchadores compiten en el Open Galicia, en Ribeira		
0.14%	2004-2-6	La voz de Galicia

Sort by: Title | Date | Publisher | Relevance

Results per page: 5

<< Previous results Next results >>

OmniPaper: The smartest European news finder

Overall Knowledge Layer prototype

Article

Bananenman opgepakt wegens smokkel illegale kruiden

Tien om te Zien is terug in het land, sidder en beeft!!

Vilvoorde ? Na herhaalde protest-acties van de Vlaamse zang -en bleit Gilde, lijkt er in het Vlaamse medialandschap terug beweging te komen. ? Het is de laatste tijd niet te doen? zegt Bananenman, vertegenwoordiger van de Vlaamse zang -en bleit Gilde. ?De overheid doet niets ter ondersteuning van het Vlaams gebleit en de media heeft ons ook gelijk een baksteen laten vallen. Enkele jaren geleden hadden we nog TOTZ, maar nu de Willy alweer een jonge deerne heeft zwanger gemaakt (nvdr. Willy Sommers is een Vlaams charmezanger die vooral jonge vrouwen charmeert na de optredens), is het programma definitief stop gezet.? De platenverkoop is volgens EMI, de platenmaatschappij die de meeste Vlaamse bleitkoppen vertegenwoordigd, de laatste jaren sterk gedaald. ?Het is allemaal de fout van die illegale downloaders? krijgen we als grootste argument te horen van de persverantwoordelijke, Marina. Het ziet er dus naar uit dat de Vlaamse nachtegalen echt moeten werken om aan de bak te komen, parochiezalen, rusthuizen, familiefeesten en zelfs jeugdhuizen zullen dus niet meer veilig zijn. U hebt het vast al wel meegemaakt, de plaatselijk Jo Vally al bleitend op een begrafenis. ?Er moet dringend iets gebeuren, zoniet volgen er hardere acties? vervolgt Bananenman. ?Wij hebben in onze laatste ledenvergadering gestemd voor verdere acties, met een overgrote meerderheid pro Zullen we her en der op de Belgische wegen al zingend in onze BMW ZXITurboBoxer II Cabrio Vlaamse Schlagers te berde brengen. En als we echt niet anders kunnen zullen we extremistische acties moeten ondernemen, we deinzen niet terug om Helmut Lotti een duet te laten zingen met La Estarella.? Als reactie hierop heeft VRT baas Jefke de lathouwer, Jos Mommen aangesteld om de haalbaarheid van een 3e zang -en bleit net Op te starten. Wordt ongetwijfeld vervolgd, en ondertussen bang afwachten!

80% NL 2004-3-22

Keywords: Banaan , Cartoon , BBC

OmniPaper: The smartest European news finder

OmniPaper

The smartest European news finder

Demo

Contents

- OmniPaper project introduction
- "Local Knowledge Layer" prototypes:
  - SOAP
  - Automatic Keyword Extraction (AKE)
  - RDF
  - Topic Maps
- Comparison of prototype performance
- "Overall Knowledge Layer" architecture
- "Overall Knowledge Layer" prototype
- → Conclusions and future work

Conclusions

- Distributed architecture:
  - + Powerful
  - + Flexible
  - + SOAP approach: minimal effort by local archive
  - - Bookkeeping (copyrights)
  - - Tension between push vs pull approach
    - → OmniPaper proposes a combined push and pull method
- Search performance:
  - Full-text searching performs better than pure semantic search with WordNet (Topic Maps)
  - Semantic query expansion seems to give better result ranking (higher precision at lower recall rates)
  - More research is needed with more elaborate test sets

Conclusions

- Advantages of EuroWordNet
  - Query expansion through concepts
  - Extended navigation possibilities
    - User can browse to related concepts
    - Query feedback
  - Multilingualism
    - Integrate WordNets in different languages
    - ILI is very powerful mechanism
- Disadvantages:
  - Not really commercially exploited and maintained: only English is up-to-date
  - Quality varies across languages
  - Too fine-grained for web of concepts (semantic differences sometimes too small)
  - Limited support of proper names
    - For news searching this is very important

Conclusions

- Query vs navigation
  - Web of concepts powerful tool for semantic query feedback
  - Query and navigation are both methods of searching and should be used in combination with each other
    - Query = starting point for search
    - Navigation:
      - Give semantic feedback on the query
      - Refine the query
      - Browse away from the query

Future work

- Testing of different search methods
  - Methods:
    - Full-text (SOAP)
    - Full-text (SOAP) with semantic query expansion (Topic Maps)
    - Keyword search (RDF)
    - Keyword weights matching (AKE)
    - Pure semantic search (Topic Maps)
  - Cross-testing with bigger CLEF dataset

Future work

- Further development of business plan
  - OmniPaper will be added service for news publishers and news brokers
  - Services:
    - Extension of their content to other providers' content using SOAP
    - Expanded search features:
      - Multilingual search support
      - Semantic query enhancements:
        - » Semantic expansion
        - » Web of concepts
        - » AKE

Contact information

- www.omnipaper.org
  - Subscribe to our newsletter !
- info@omnipaper.org
- Project coordinators:
  - Jan.engelen@esat.kuleuven.ac.be
  - Bert.paepen@esat.kuleuven.ac.be
- Technical director:
  - schranz@pressetext.at