



Prod-Trees Demonstrator Architecture



Bernard Valentin

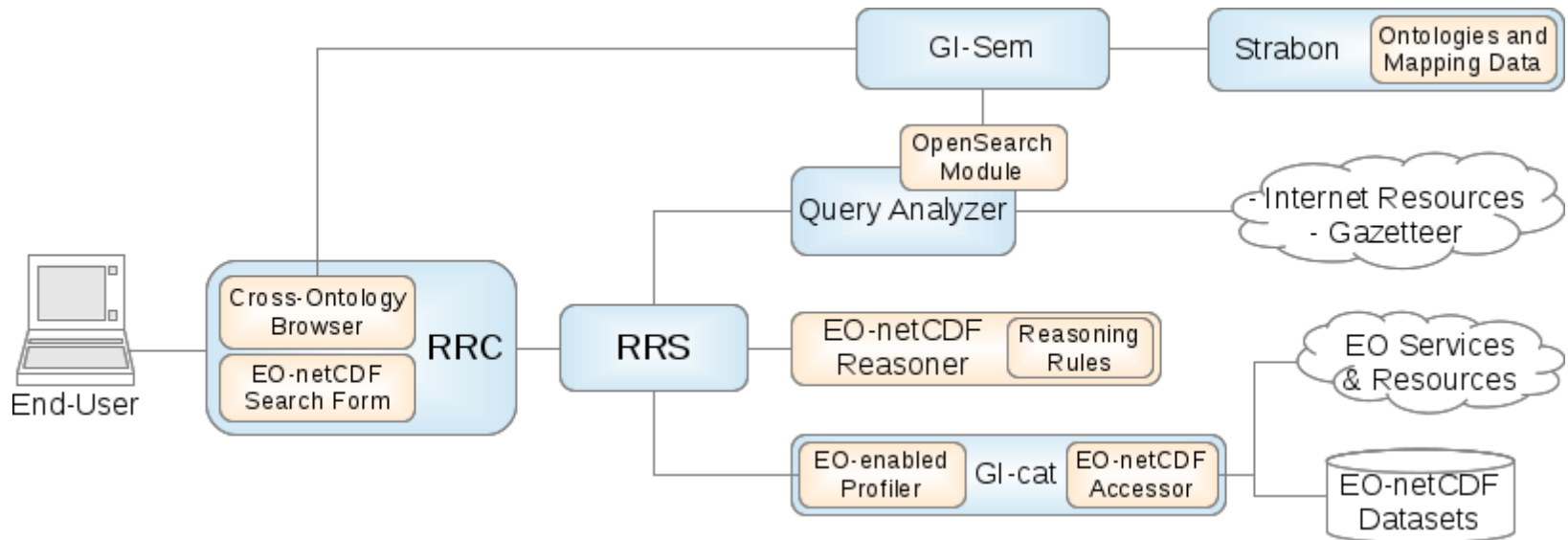
Outline

2

- Prototype Demonstrator Architecture
- User Query Lifecycle
- User Query Disambiguation

Prototype Architecture

3



Example Query Lifecycle

Interpreted Query

agriculture »»» Ontology
 Concept: production of food and goods through farming and forestry.

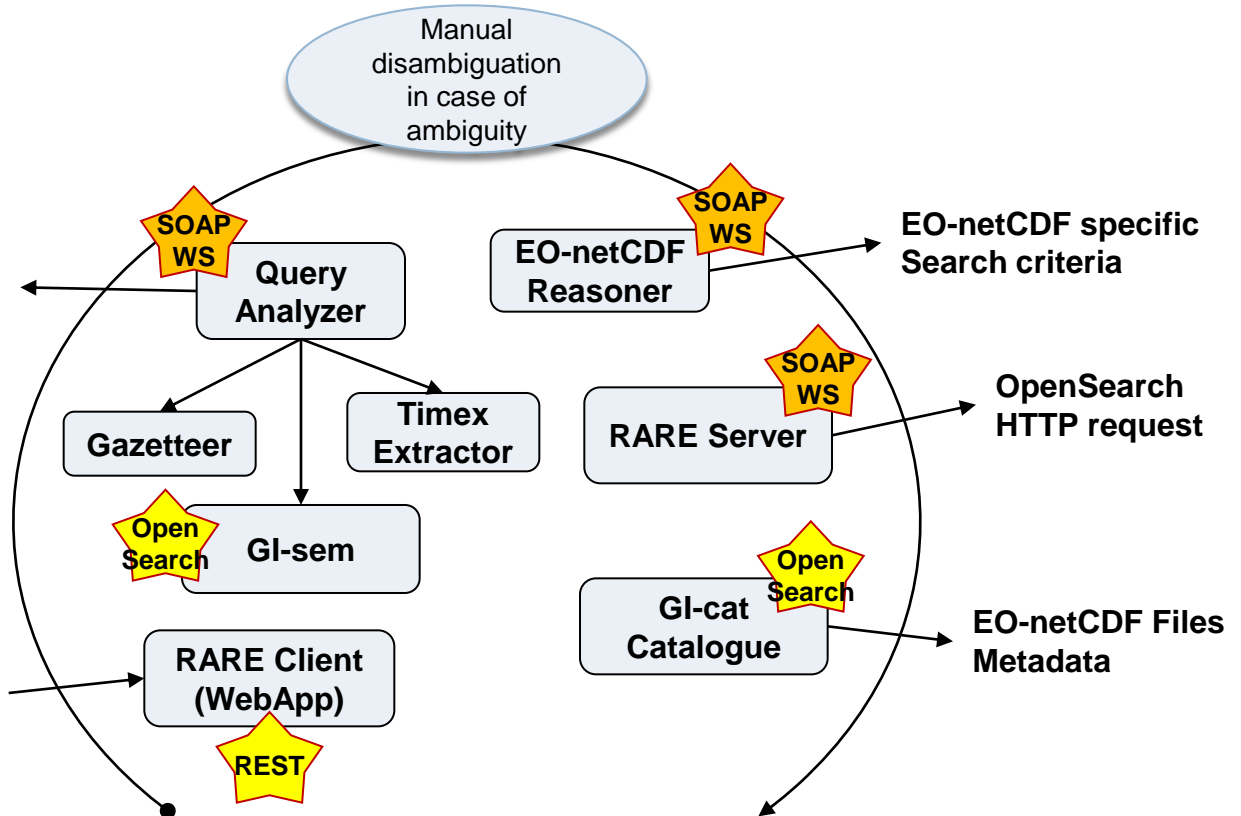
brussels »»» Toponym:
 capital of Belgium (GeoNames)
 »»» bounding box

2012 »»» Time Expression:
 between 1-Jan-2012 and 31-Dec-2012 »»» time range

Search terms:

agriculture **brussels** **2012**
 AT GT TE

AT: Application Term
GT: Geographical Term / Toponym
TE: Time Expression



User Query Disambiguation

5

- Present abilities
 - Application Term recognition: concept name recognition for the GSCDA ontology
 - Geographical Term / Toponym recognition using a geographical gazetteer (GeoNames + Google GeoCoding)
 - Common knowledge: the common English (Princeton WordNet)
 - Domain-specific terms recognition: Lucene-based lookup service presently configured with the CEOS EO Handbook (<http://www.eohandbook.com>)
 - Named Entity Recognition (DBpedia Spotlight)
 - Temporal Expression Recognition

Query Analyzer API

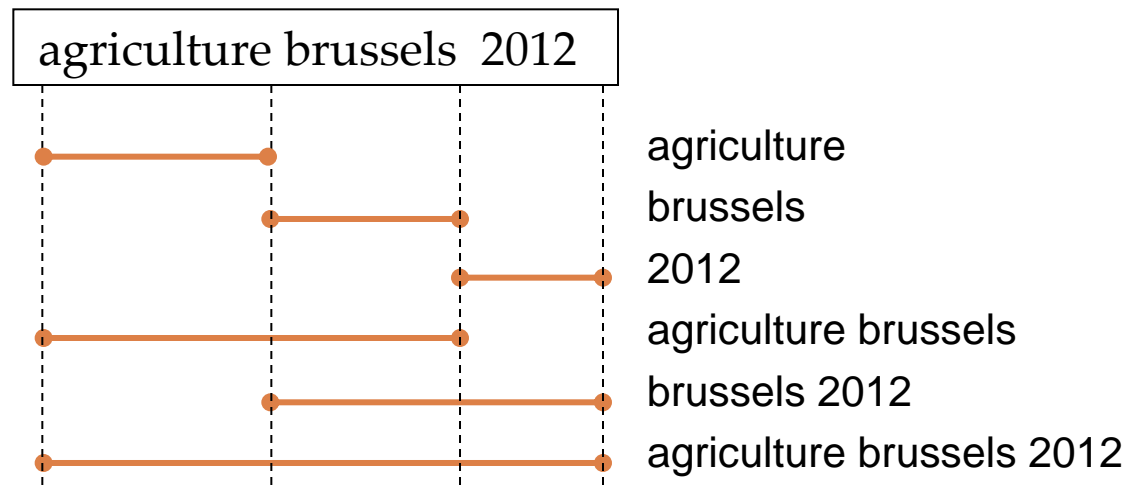
6

- The QA is implemented as a Java Web Service
- Several dependencies: GATE, Ontology Service (GI-sem), WordNet, GeoNames
- WS Operation: `getQueryInterpretation(textual_query)`
E.g. `getQueryInterpretation("agriculture brussels 2012")`
- Returned `NormalizedQuery` (XML document) contains disambiguation data obtained from the QA modules: Application Terms (OTS), WSD (WordNet), Toponyms (GeoNames), Temporal Expressions
- A client library is available to convert the module-oriented XML document into a token-oriented JSON document: `DisambiguationData`
- The JSON document contains the preferred interpretation

User Query Disambiguation

7

- The Query Analyzer looks for senses for each term and combination of adjacent terms



User Query Disambiguation

8

- The Query Analyzer looks for senses for each term and combination of adjacent terms and returns a `NormalizedQuery`, transformed into a `DisambiguationData` document:

```
{"DisambiguationData": {  
  "Interpretation": [["t0", "sid1"],["t2", "sid1"]],  
  "t0":["<token-1>", ["sid1", "sid2", "sid3"],  
    {"sid1":{<token-1-sense-1-data>}},  
    {"sid2":{<token-1-sense-2-data>}},  
    {"sid3":{<token-1-sense-3-data>}}  
  ],  
  "t2":["<token-2>", ["sid1", "sid2"],  
    {"sid1":{<token-2-sense-1-data>}},  
    {"sid2":{<token-2-sense-2-data>}}  
  ],  
  "t0 t2":["<token-1> <token-2>", ["sid1"],  
    {"sid1":{<token-1+2-sense-1-data>}}  
  ],  
  "NormalizedQuery": "<NormalizedQuery-XML>"  
}}
```


User Query Disambiguation

9

- If there is no ambiguity
 - ▣ the most probable meaning is displayed to the user

Search Terms

Insert application term

Disambiguated Query

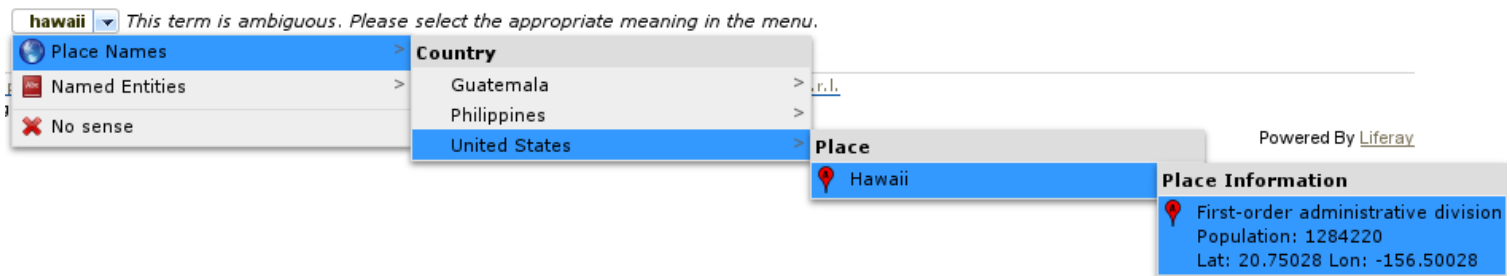
london **London** is a capital of a political entity located in **United Kingdom** (GeoNames).
2002 from 1 January to 31 December 2002.

- ▣ the search for EO-netCDF products starts immediately

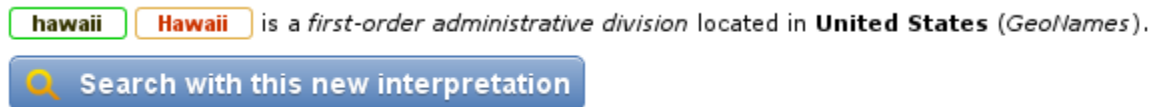
User Query Disambiguation

10

- Otherwise,
 - ▣ the user is invited to select a sense in contextual menus



- ▣ the search for EO products must be triggered by the user (button click)



Thank you