

Natural Language Interfaces to Ontologies

Danica Damljanović
danica@dcs.shef.ac.uk



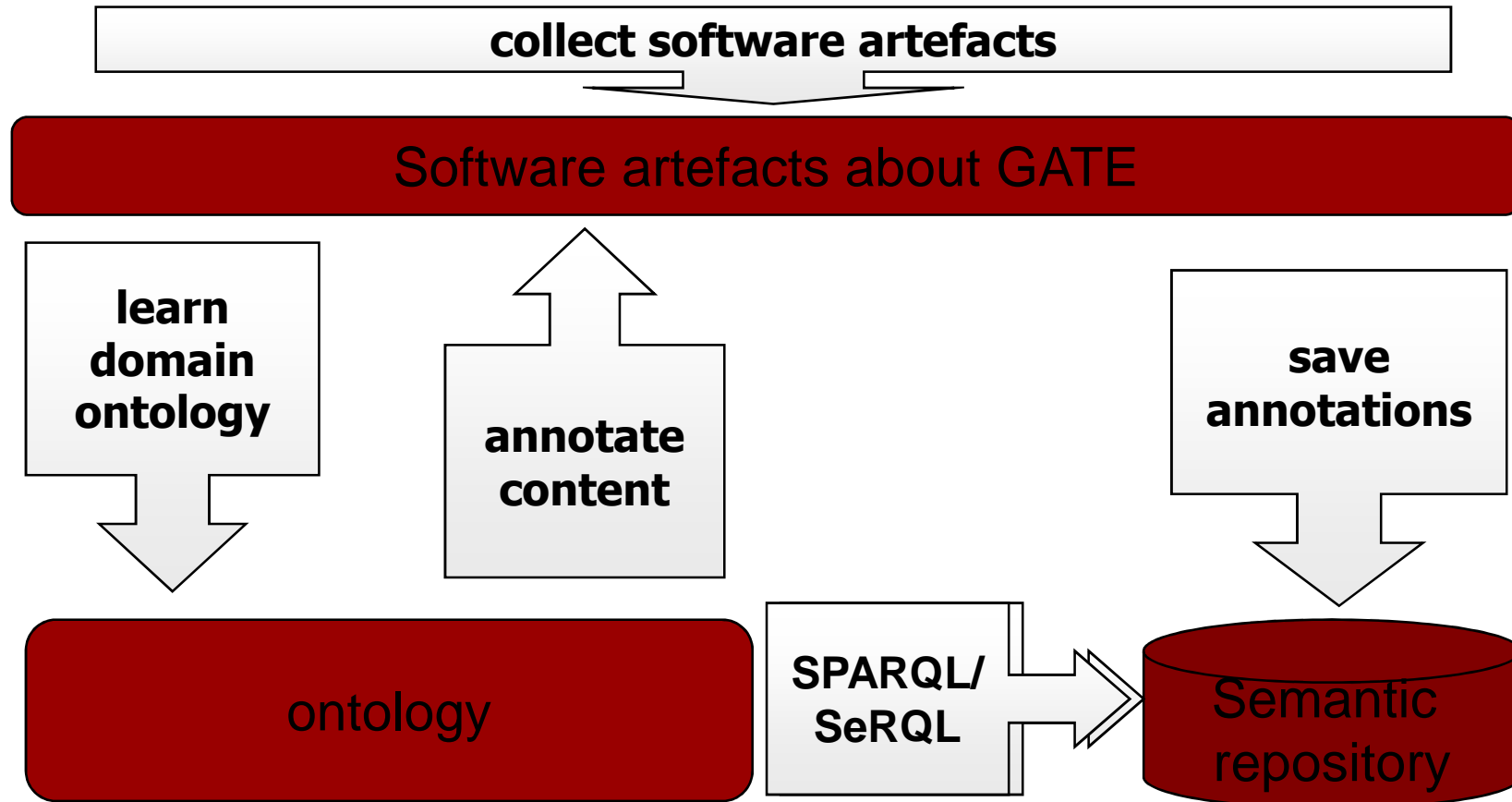
Sponsored by

- Transitioning Applications to Ontologies:
 - www.tao-project.eu





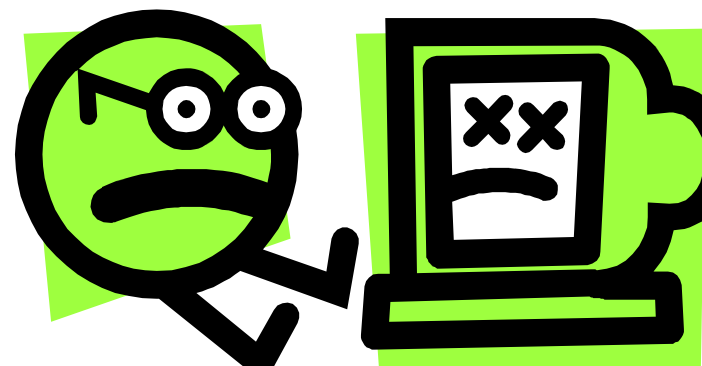
GATE case study in TAO project





Query languages

- SPARQL/SeRQL:
 - complex syntax: not easy to learn
 - writing queries is error-prone task
 - requires understanding of Semantic Web technologies





Objective

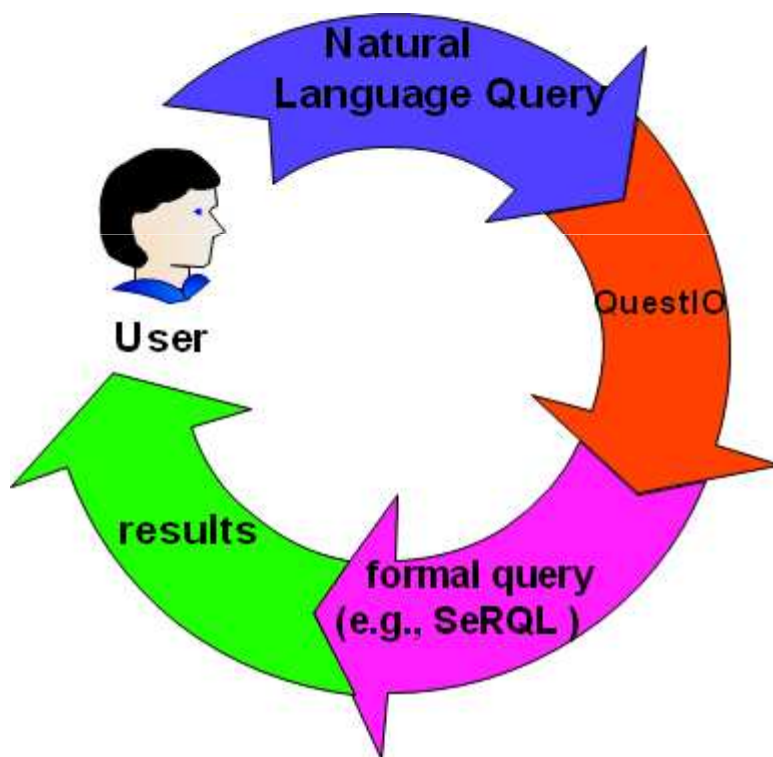


Allow domain experts to query knowledge in RDF/OWL format in a user-friendly manner





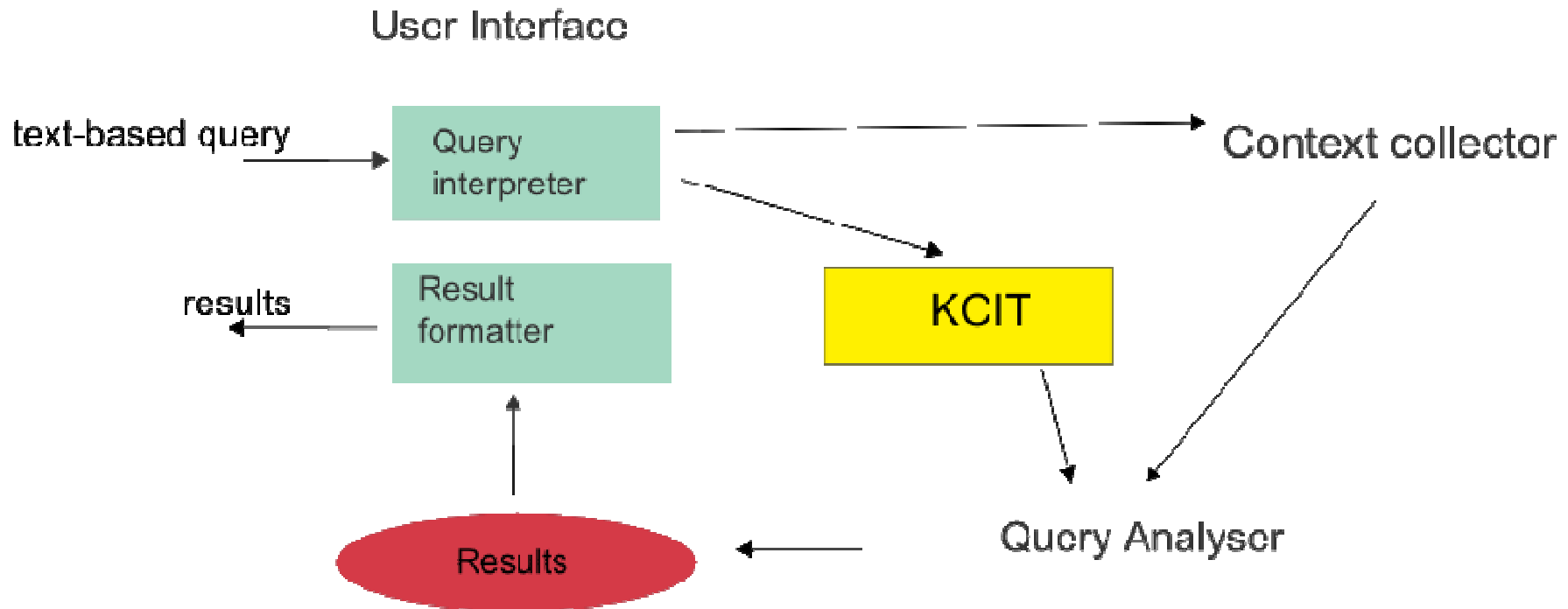
Question-based Interface to Ontologies



- Robust
- Zero customisation!
- Easy to use: no training for the user.
 - Deal with incorrectly formulated queries
 - Accept queries of any length and form.

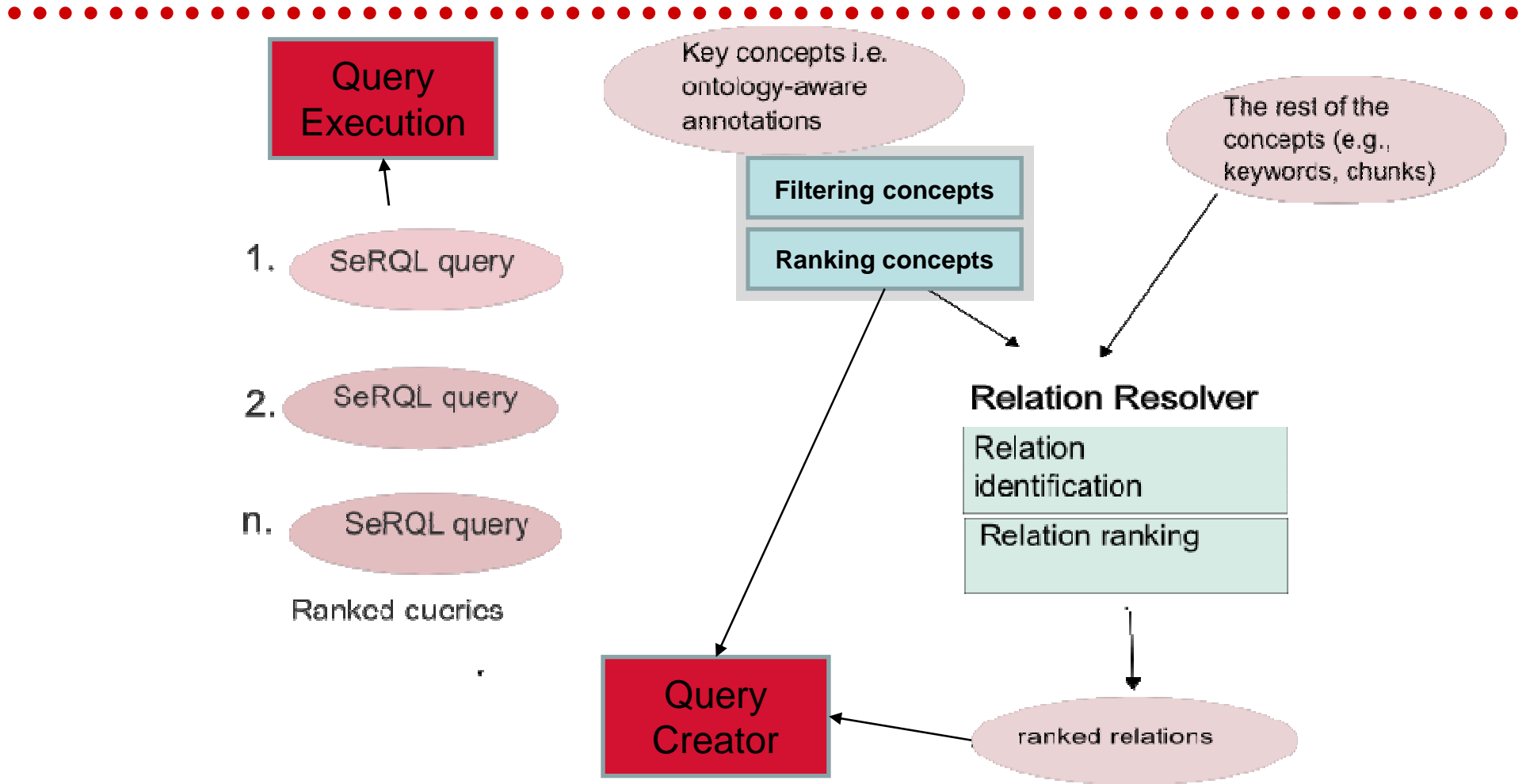


QuestIO component diagram

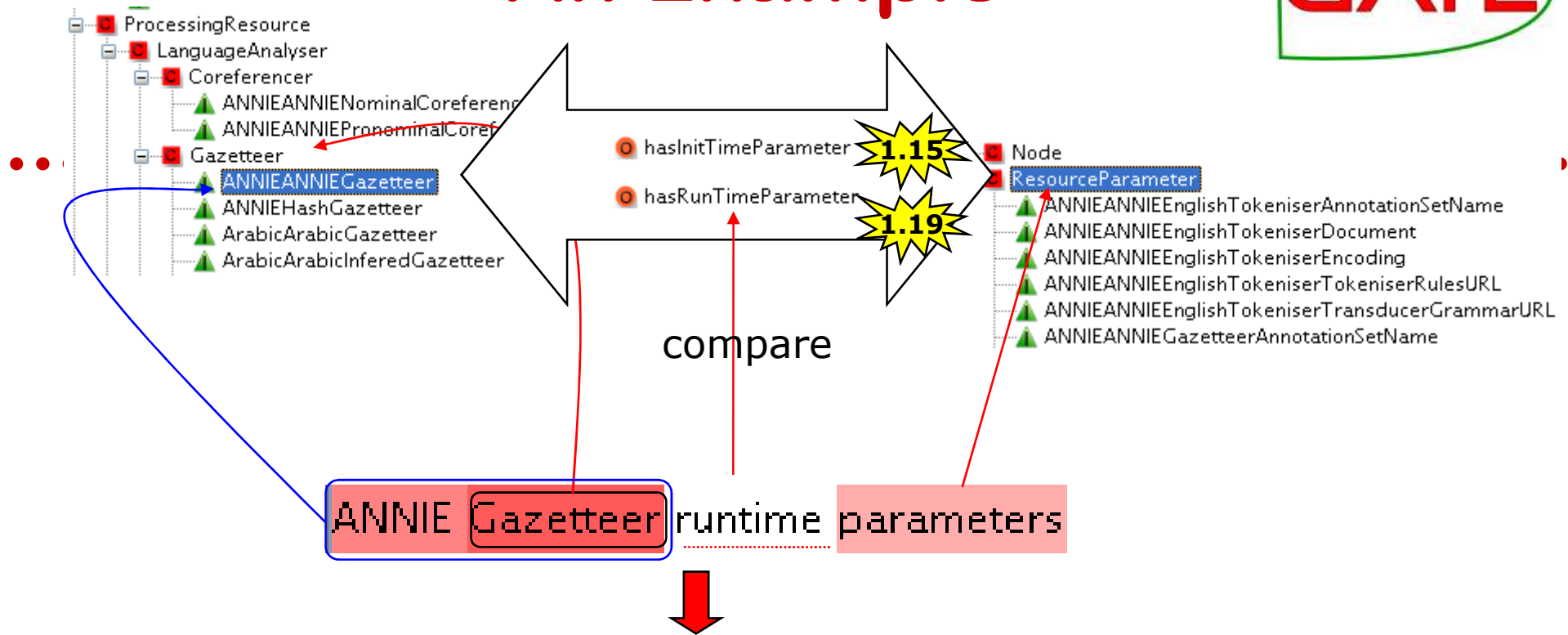




NL --> SeRQL query



An Example



Result:

- ANNIE Gazetteer --> hasRunTimeParameter --> document
- ANNIE Gazetteer --> hasRunTimeParameter --> ANNIEANNIEGazetteerAnnotationSetName
- ANNIE Gazetteer --> hasRunTimeParameter --> ANNIEANNIEGazetteerWholeWordsOnly
- ANNIE Gazetteer --> hasRunTimeParameter --> longestMatchOnly



Hands on: closer look at the pipeline

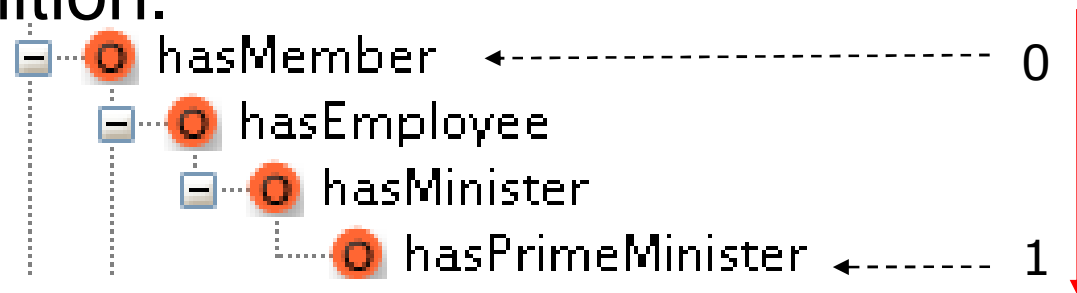
1. Download: gate.ac.uk/sale/dd/questio.zip to **U:/gate** and unzip to the same folder
2. Right click on **Applications** and **Restore application from file**
3. Select **gate.xgapp** from the unzipped questio folder
 - Double click on *query* and type in “*PRs in ANNIC*”, click on *Message* tab to see results
4. To take home: a set of sample questions and 2 sample applications covering software engineering (GATE) and tourism (Travel Guides) domains, see README file

Scoring relations



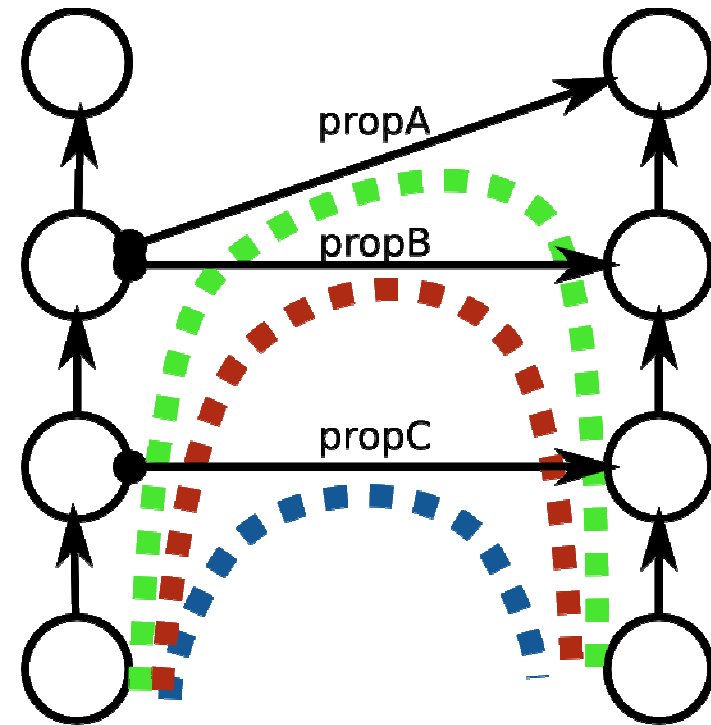
We combine three types of scores:

- *similarity score* - using Levenshtein similarity metrics we compare input string from the user with the relevant ontology resource
- *specificity score* is based on the subproperty relation in the ontology definition.



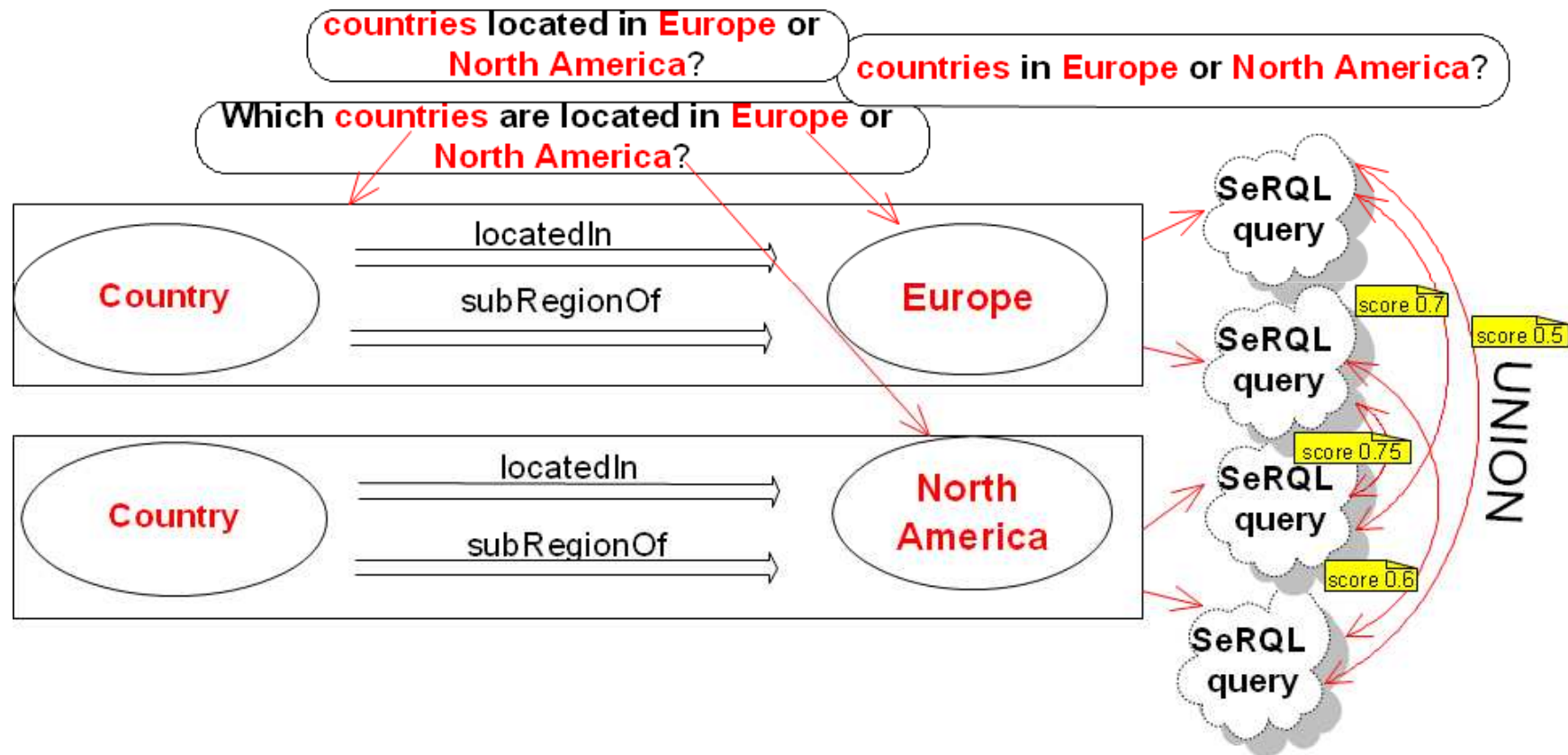
Scoring relations (II)

- *distance score is inferring an implicit specificity of a property based on the level of the classes that are used as its domain and range.*



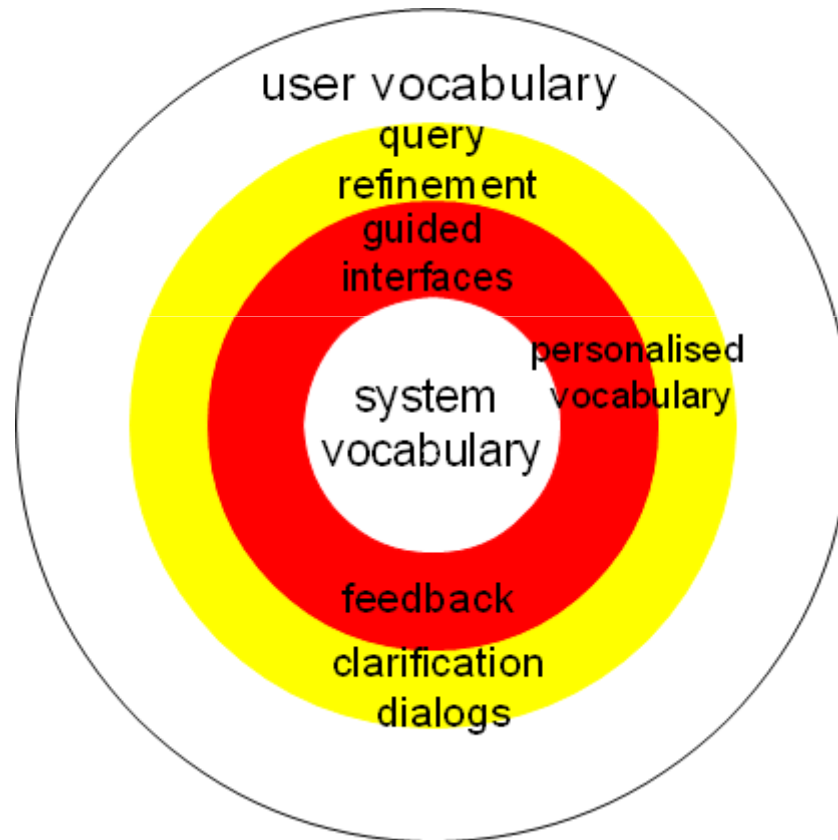


Disjunction/conjunction queries





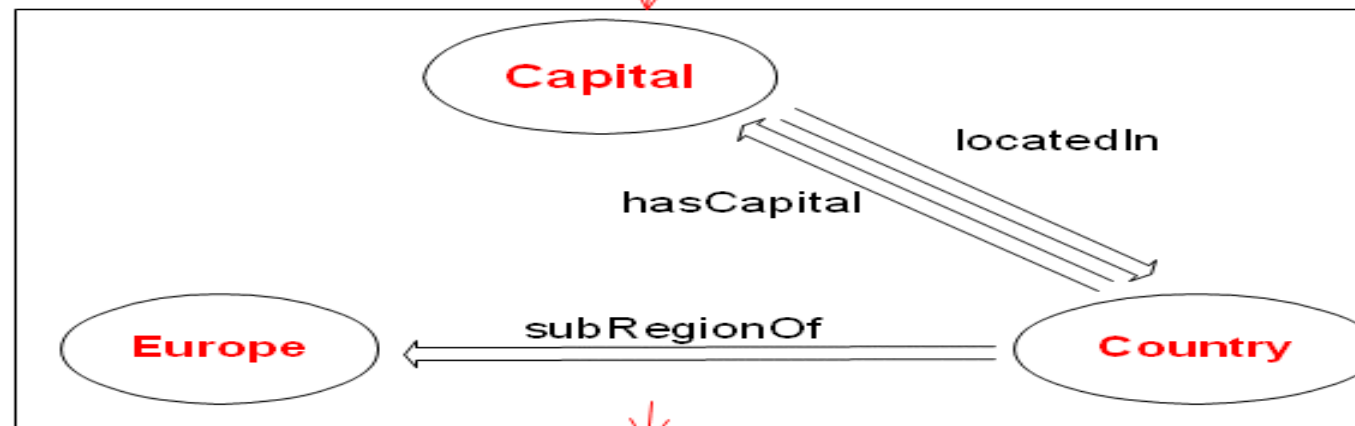
Can the user and the system speak the same language?



capitals of countries in Europe

capitals of countries located in Europe

What are the capitals of countries located in Europe?



```
select c0,"[inverseProperty]", p1, c3, p4, i6
from {c0} rdf:type {<http://proton.semanticweb.org/2005/04/protonu#Capital>},
{c3} p1 {c0}, {c3} rdf:type {<http://proton.semanticweb.org/2005/04/protonu#Country>},
{c3} p4 {i6}, {i6} rdf:type {<http://proton.semanticweb.org/2005/04/protonu#Continent>}
where p1=<http://proton.semanticweb.org/2005/04/protonu#hasCapital> and
p4=<http://proton.semanticweb.org/2005/04/protonu#subRegionOf> and
i6=<http://www.ontotext.com/kim/2005/04/wkb#Continent_T.4>
```

score: 1.55

```
select c0, p1, c3, p4, i6
from {c0} rdf:type {<http://proton.semanticweb.org/2005/04/protonu#Capital>},
{c0} p1 {c3}, {c3} rdf:type {<http://proton.semanticweb.org/2005/04/protonu#Country>},
{c3} p4 {i6}, {i6} rdf:type {<http://proton.semanticweb.org/2005/04/protonu#Continent>}
where p1=<http://proton.semanticweb.org/2005/04/protonu#locatedIn> and
p4=<http://proton.semanticweb.org/2005/04/protonu#subRegionOf> and
i6=<http://www.ontotext.com/kim/2005/04/wkb#Continent_T.4>
```

score: 0.85

executing the SeRQL query with the highest score

Paris --> [inverseProperty] hasCapital --> French Republic --> subRegionOf --> Europe
Budapest --> [inverseProperty] hasCapital --> Republic of Hungary --> subRegionOf --> Europe
Prague --> [inverseProperty] hasCapital --> Czech Republic --> subRegionOf --> Europe
Helsingfors --> [inverseProperty] hasCapital --> Republic of Finland --> subRegionOf --> Europe
Belgrade --> [inverseProperty] hasCapital --> Serbia and Montenegro --> subRegionOf --> Europe
Bern --> [inverseProperty] hasCapital --> Swiss Confederation --> subRegionOf --> Europe
Copenhagen --> [inverseProperty] hasCapital --> Kingdom of Denmark --> subRegionOf --> Europe
Athens --> [inverseProperty] hasCapital --> Greece --> subRegionOf --> Europe
.....



FREyA (Feedback, Refinement, Extended Vocabulary Aggregator)

Query:

Identified context	Our score [0-100]
capital has capital country sub region of europe (continent)	0.13 <input checked="" type="radio"/>
capital has capital country located in europe (continent)	0.13 <input type="radio"/>
capital part of country sub sector of europe (continent)	0.12 <input type="radio"/>
capital sub region of country registered in europe (continent)	0.12 <input type="radio"/>
capital sub region of country established in europe (continent)	0.12 <input type="radio"/>

done

```
graph LR; A["country (49)"] -- has capital --> B["capital (49)"]; A -- sub region of --> C["europe (1)"]
```

country (49)

- Czech Republic
- Republic of Malta
- jersey
- Principality of Monaco
- Former Yugoslav Republic of Macedonia
- Republic of Estonia
- Isle of Man
- Italian Republic
- Holy See
- Republic of Ireland



Demo

-
- <http://gate.ac.uk/freya>



Hands on and evaluation

- Open Firefox:
 - Go to: <http://gate.ac.uk/freya>
- Follow instructions on the handouts
- Tasks 1a, 2a, 3a about GATE
- Tasks 1b, 2b, 3b about geography of the United States*
- Do a **or** b, not both



*Ontology courtesy of The University of Zurich and The University of Texas



Evaluation on coverage and correctness



- 36 questions extracted from GATE list
- 22 out of 36 questions were *answerable* (the answer was in the knowledge base):
 - 12 *correctly answered* (54.5%)
 - 6 *with partially corrected answer* (27.3%)
 - system *failed* to create a SeRQL query or created a wrong one for 4 questions (18.2%)
- Total score:
 - **68%** correctly answered
 - 32% did not answer at all or did not answer correctly

Comparison with AquaLog

- Removed 6 questions not supported by AquaLog:
 - “What are the run parameters of POS Tagger and Sentence splitter?”
 - “Does GATE have a coreference resolution component (PR)?”
 - “How many. . . .”
 - “I cannot get Wordnet plugin to work“.

	QuestIO	Correct	Aqualog	Correct
correct	9 (56.25%)	56.25%	5 (31.25%)	31.25%
c. correct	0	(0%)	3 (18.75%)	18.75%
p. correct	5 (31.25%)	15.63%	3 (18.75%)	9.35%
failed	2 (12.5%)	0%	5 (31.25%)	0%
		71.88%		59.35%

Evaluation on scalability and portability



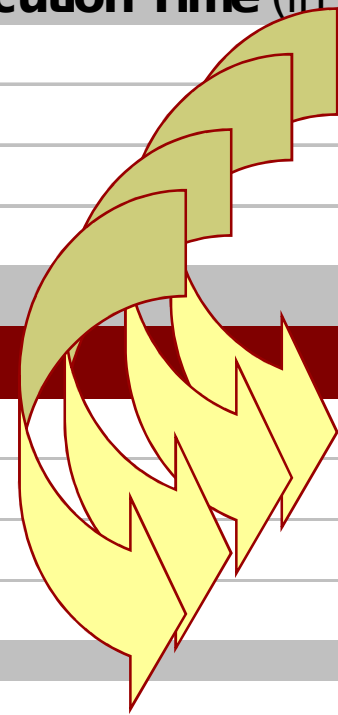
- Ontologies have not been customised or changed prior using with QuestIO!
 - GATE ontology: <http://gate.ac.uk/ns/gate-kb>
 - Travel ontology: <http://goodoldai.org/ns/tgproton.owl>

	GATE kb	Travel kb
Classes	42	318
Object Properties	23	86
Instances	594	2790
Total size	659	3194
Initialisation time	16 seconds	22 seconds

Scalability and portability: execution time



GATE knowledge base	
	Execution Time (in seconds)
what are the parameters of pos tagger?	0.17
pos tagger parameters?	0.17
pos tagger runtime parameters?	0.105
parameters of prs included in ANNIC or ANNIE	0.34
Travel knowledge base	
which countries are located in europe?	0.22
capitals of countries	0.23
is london capital of any country?	0.2273
capitals of countries in asia or europe	0.4



Thank you!

-
- Questions?

