



Crosslingual retrieval in Semantic Web

Cristina Vertan,

University of Hamburg • Informatics Department
Natural Language Systems Group

WWW: <http://nats-www.informatik.uni-hamburg.de/~cri/>

E-Mail: cri@nats.informatik.uni-hamburg.de

Aims of the system

- To allow the user to express a query in natural language
- To ensure that relevant information from the query is interpreted and processed by the system
- To allow retrieval of corresponding information from texts in several languages specified by the user

Contents

- System architecture
- Controlling the natural language input
- Feature extraction
- Ontology and multilinguality
- Semantic Web annotation for text retrieval
- Generation of the answer

What are we doing? - General Ideas

- No general NLP system, because ...
 - It is not feasible with any grammar, any domain, any inference type
- How to restrict the problem?
 - Hybrid system with pictorial and linguistic interface
 - Tailored for one domain
 - Only a few syntactic and semantic patterns
 - only one pragmatic setting
- What we gain:
 - It's more specific than keyword search
 - Independent of suppliers' keywords
 - Faster than hierarchical clicking and selection
 - Cognitive approach
 - Multimodal processing

Welcome to Romania On-Line

Use our interactive map, or ask our expert



[General Info](#)

[Landscape](#)

[Regions](#)

[Round-trips](#)

.....



Ask me...

12.09.2005

SW Technologies for MT · MTSummit X

5



Transylvania



[Cities](#)

[History](#)

[Round trips](#)

.....



Ask me...

12.09.2005

SW Technologies for MT · MTSummit X

6



Transylvania -Cities



Hunedoara



Sighisoara



Ask me...



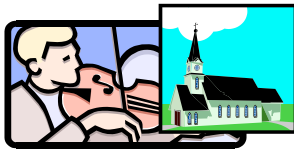
Sibiu



Brasov



Ask me...



Culture



Travel



Sport



Which

How long

Where

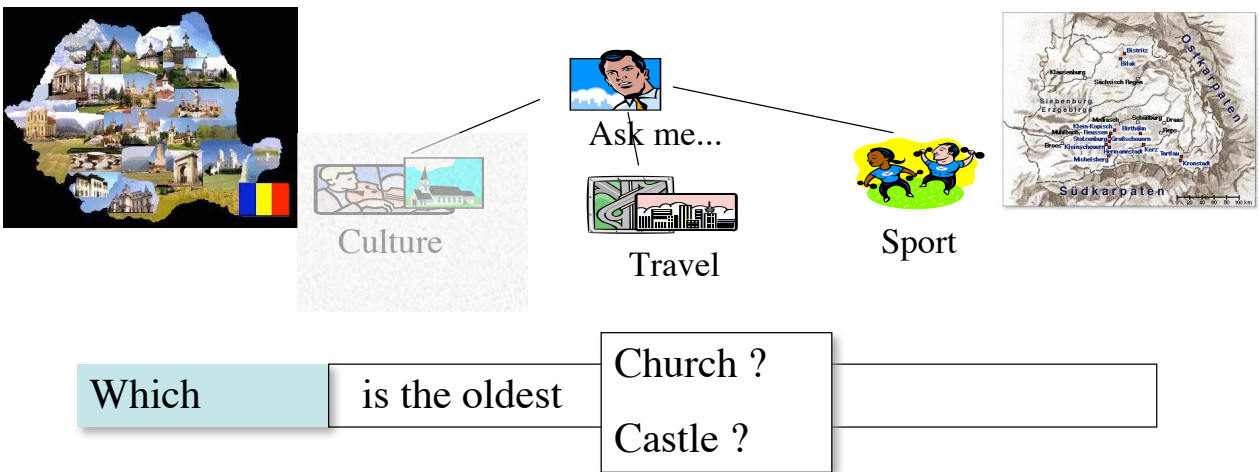
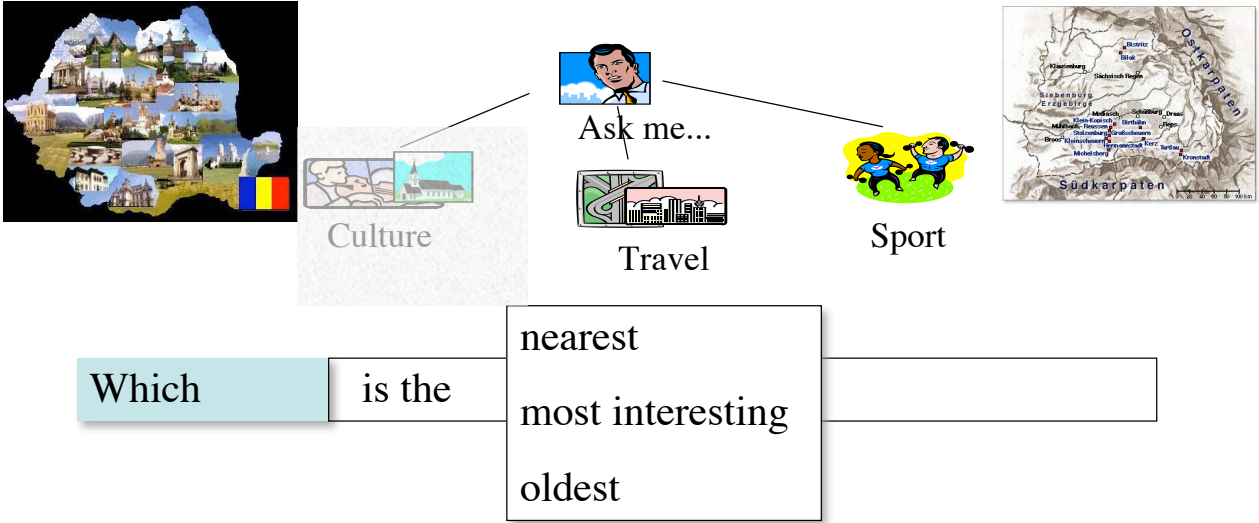
Can I

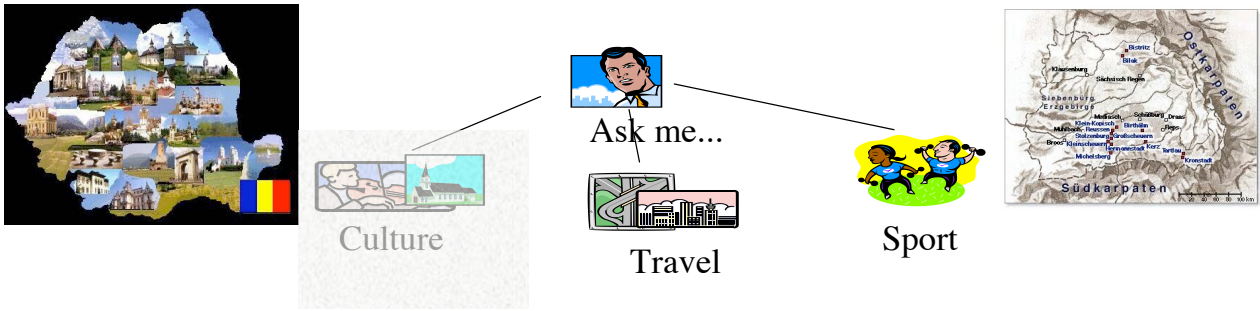
How much



Which

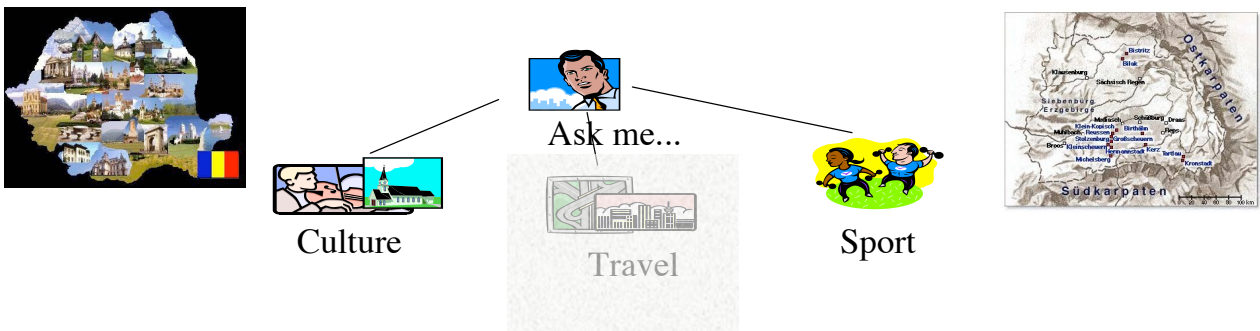
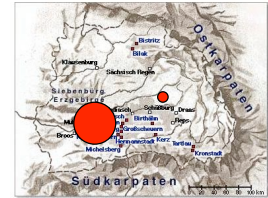
is the
are the





Which is the oldest castle ?

- The oldest entirely preserved is in [Hunedoara](#)
see also [circuit 2](#)
- Ruines of olderer castles are in



Which

How long

Where

Can I

How much



Culture



Ask me...



Travel



Sport



Can I

travel
drive
fly to



Culture



Ask me...



Travel



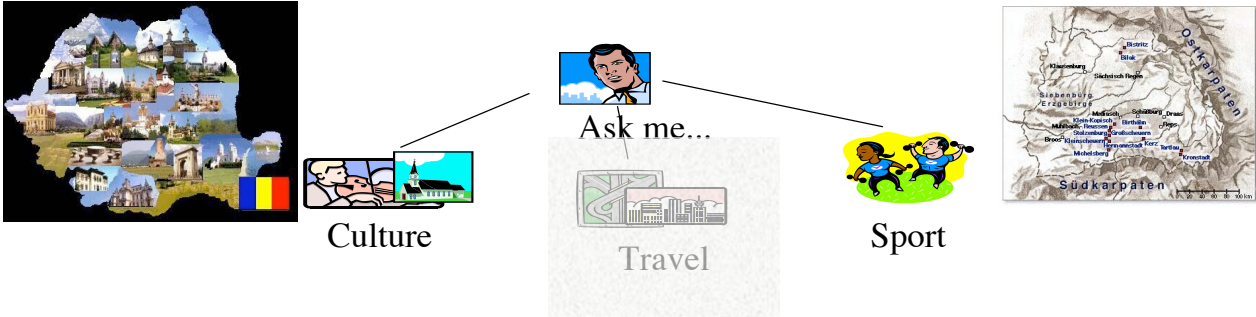
Sport



Can I

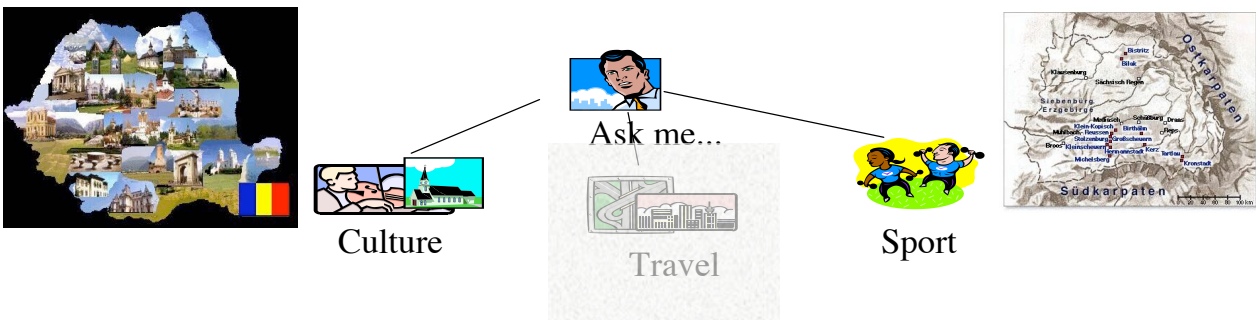
fly

Sibiu
Sighisoara
Hunedoara
.....



Can I

fly to Hunedoara

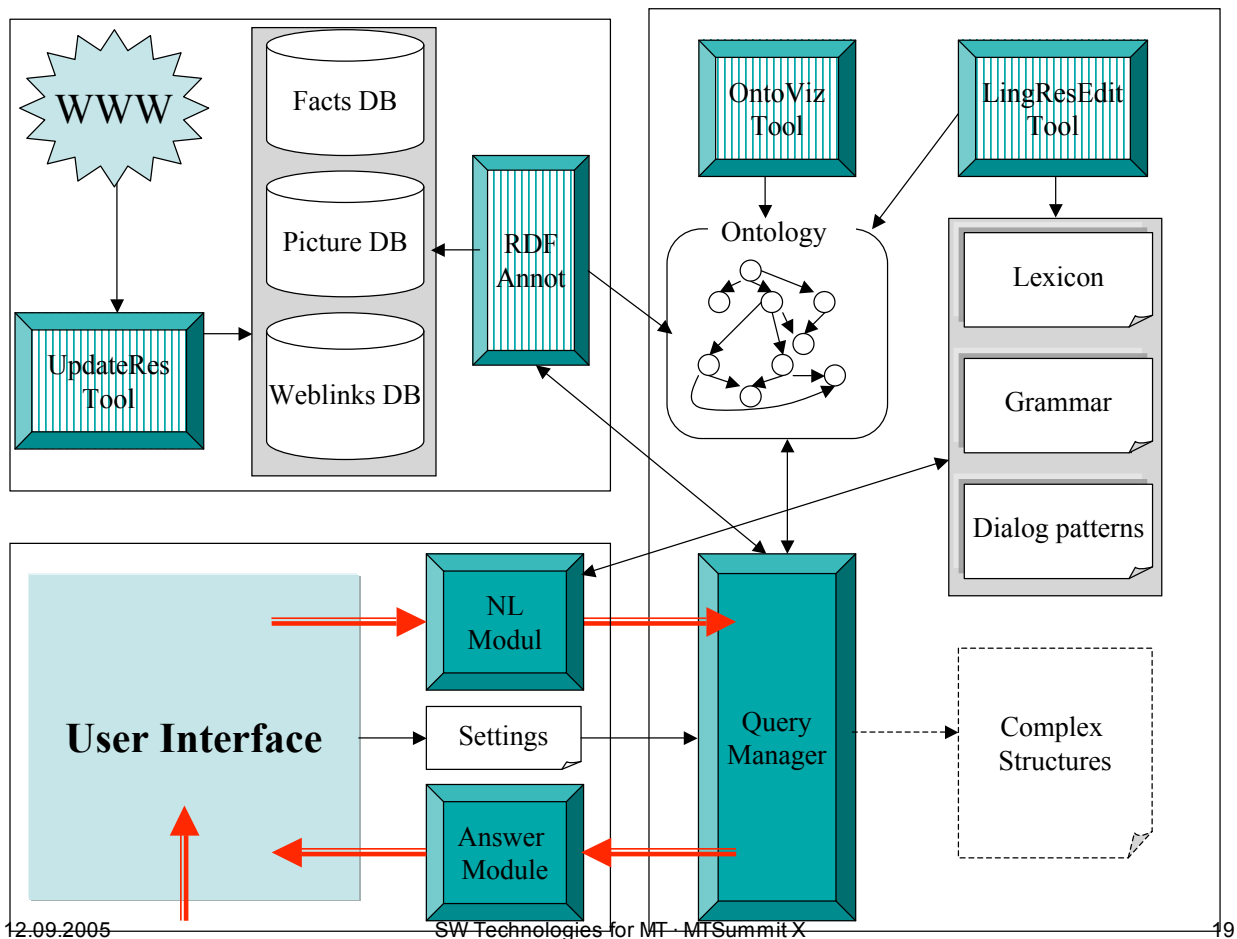


Can I

fly to Huunedoara

•No, in Transylvania you can fly to [Sibiu](#), [Cluj](#) and [Timisoara](#)





German/English Corpus (Facts and Weblinks DB) - example

Das Rothenburg Siebenbürgens
 Das historische Zentrum von Sighisoara (Schässburg) wurde von der UNESCO auf die Liste des Weltkulturerbes gesetzt. Und das zu Recht. In der Mitte steht wie einst die mächtige Burg, die besterhaltene Siebenbürgens, die im 12. und 13. Jahrhundert auf den Ruinen eines römischen Kastells errichtet wurde. Das auffälligste und sehenswerteste Gebäude ist der 64 m hohe Stundturm .

.....
<http://reisen.transylvaniatravel.net/s-taedteurlaub/reise311.html>

Sighisoara
 Strategically overlooking a valley full of picturesque villages boasting fantastic fortress churches, Sighisoara, with its small but beautifully preserved medieval core seems to be set for a prosperous future in tourism. But where are the tourists? While an attractive town like this would certainly be swamped daily by coachloads of visitors if it were in Hungary or Austria, Sighisoara can often be eerily quiet and deserted, with the notable exception of the yearly Medieval Folk Festival.

.....
<http://www.inyourpocket.com/romania/sighisoara/en/>

What to insert in the Lexicon

- For each word or expression (can I, is there, etc.) define:
 - Type of sublexicon (general, culture, sport, travel)
 - Semantical class
 - Semantical restrictions
- Example of entry:

```
<entry word=„climb“>
  <lexicon > sport </lexicon>
  <sem_class> <action_verb></sem_class>
  <sem_restriction> [empty, region, lanscape_high]
</sem_restriction>
</entry>
```

Display lexical equivalence classes ("fillers") in a sequence

Display lexical equivalence classes ("fillers")
in a sequence

- According to :
 - Semantic sequence patterns of sentences
 - Semantic classes of lexemes
 - Semantic restrictions of lexemes

Example:

- Pattern : Loc_question | aux_struct | action_verb | action_region
- Classes:
 - Loc_question {where, from where, to where}
 - Aux_struct {can I, can we}
 - Action_verb {climb, ski, swim}
 - Action_region {Transylvania, Black Sea}

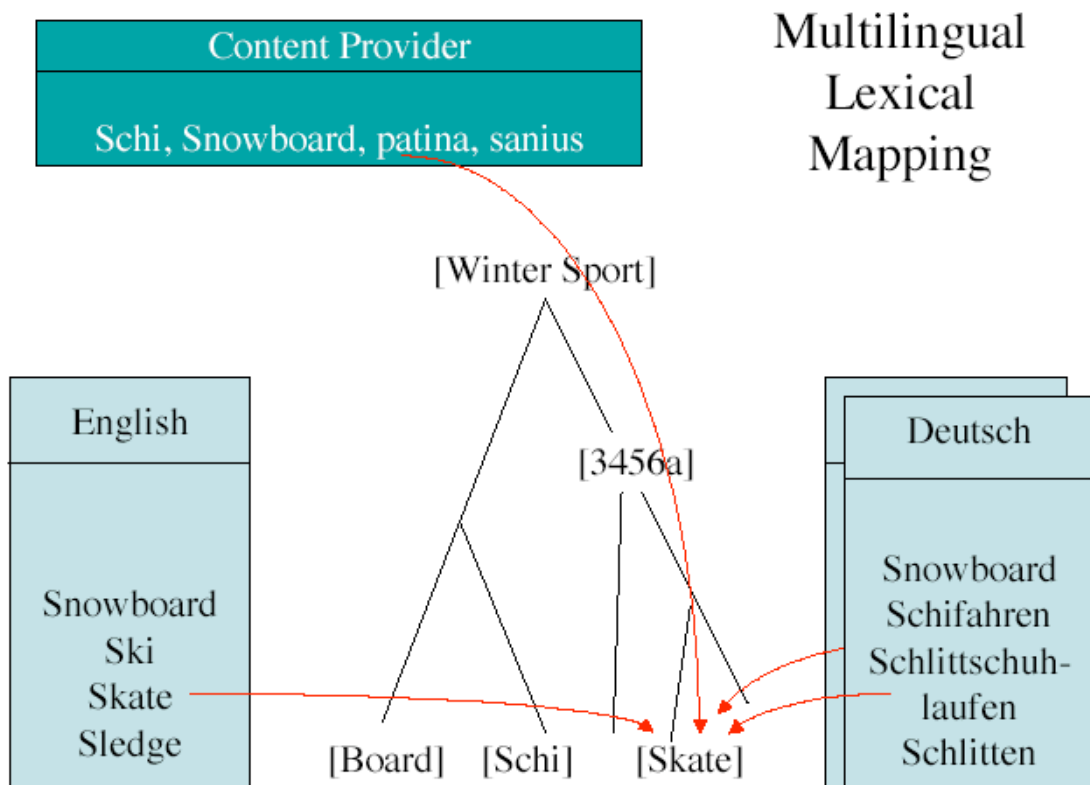
Feature extraction

- Loc_question | aux_struct | action_verb | action_area
- (Var_location , action_verb, action_area)
- A search with the values through the RDF-annotated DB: [action_verb][location]
- And all locations will be extracted.
- Filter by action_area with the help of ontologies

12.09.2005

SW Technologies for MT · MTSummit X

23

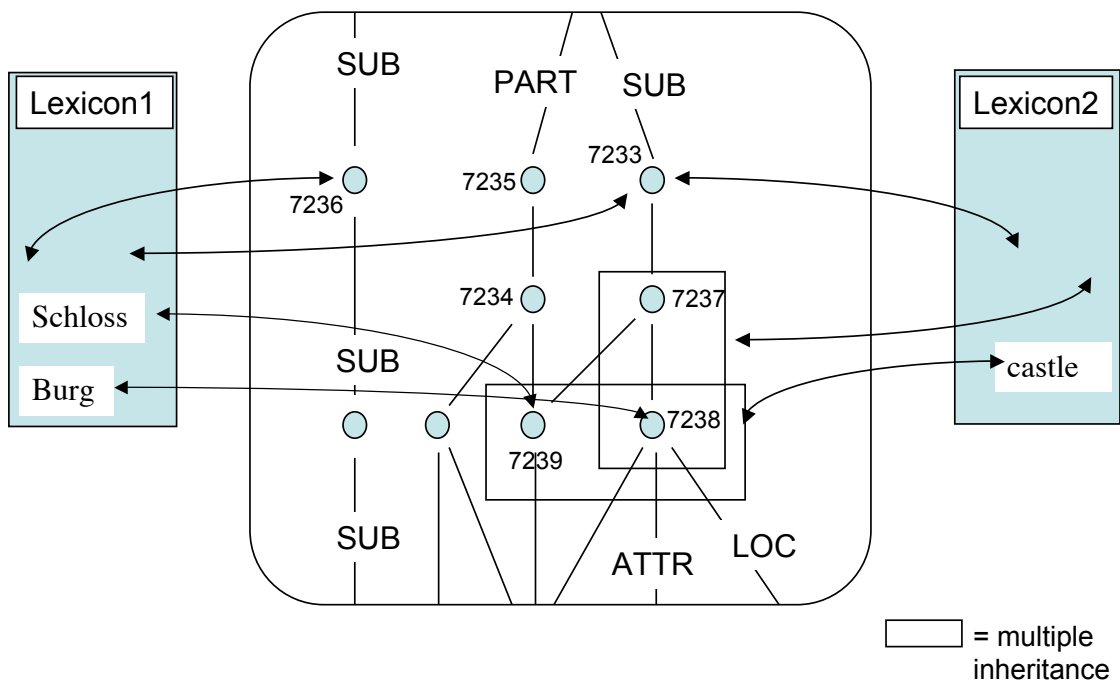


12.09.2005

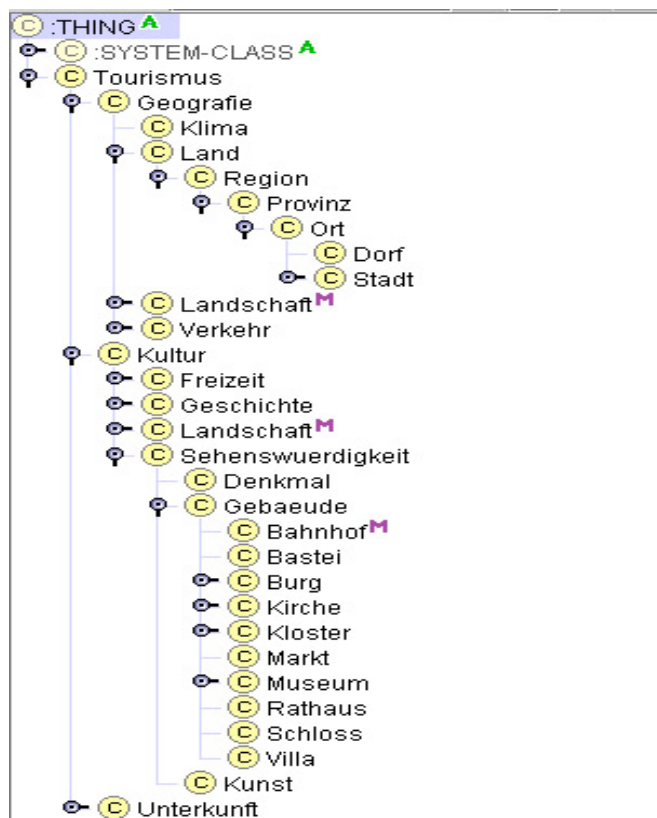
SW Technologies for MT · MTSummit X

24

Multilingual lexical mapping



Ontology - Excerpt



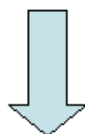
RDFS Ontology-Encoding

- `<rdfs:Class rdf:ID="Bahnhof"`
- `rdfs:label="Bahnhof">`
- `<rdfs:subClassOf rdf:resource="Eisenbahn"/>`
- `<rdfs:subClassOf rdf:resource="Gebaeude"/>`
- `<rdfs:seeAlso>Railroadstation</rdfs:seeAlso>`
- `</rdfs:Class>`

Problem with RDFS: no way to express partial synonymy

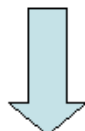
DB Mapping

Where can I skate in Transylvania ?



Search question structure

[LOC_Question] [Aux_structure] [Action_verb] [action_area]



Extract relevant information for this structure

Search: LOC

Given: [Action_verb] = "skate"

Lexical Content mapping

Given: [Action_verb]= skate and [Action_area] = Transylvania

Search: **detailed location**

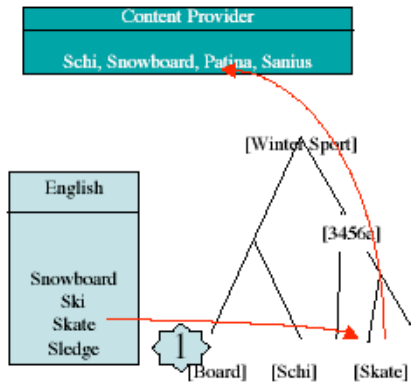


1 Retrieve the ontology concept for the English word Skate

Skate → [Skate]

2 Retrieve the Content provider's word for the concept [Skate]

[Skate] → patina

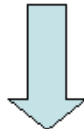


Content mapping



Retrieve all locations from the content provider where

- Action = patina and
- LOC = a city or town

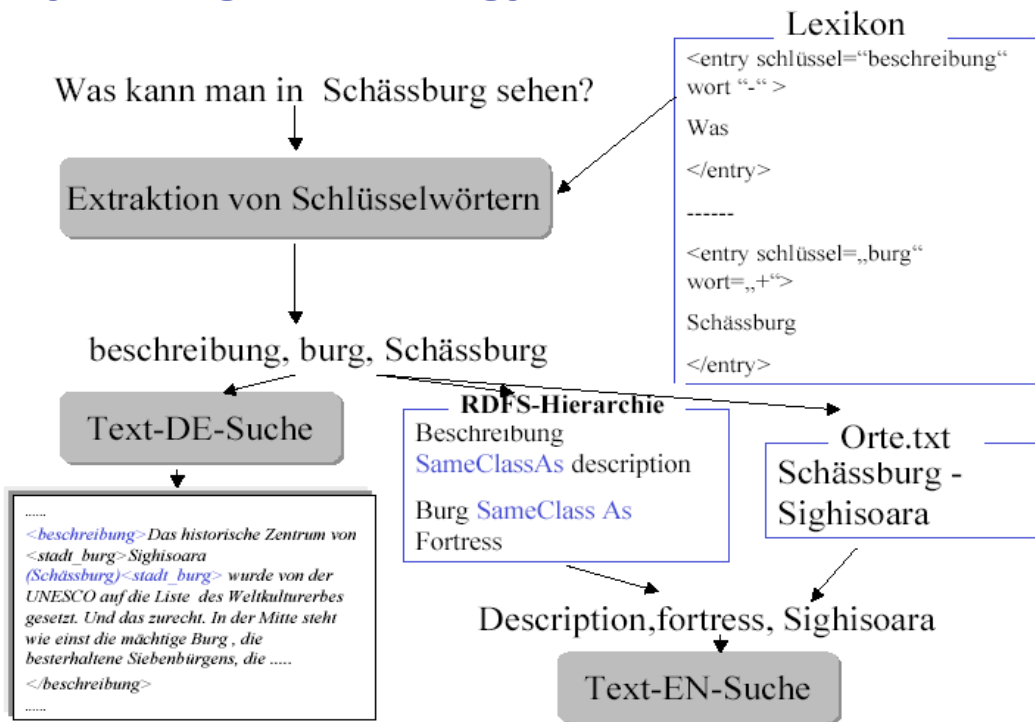


RDF-encoded

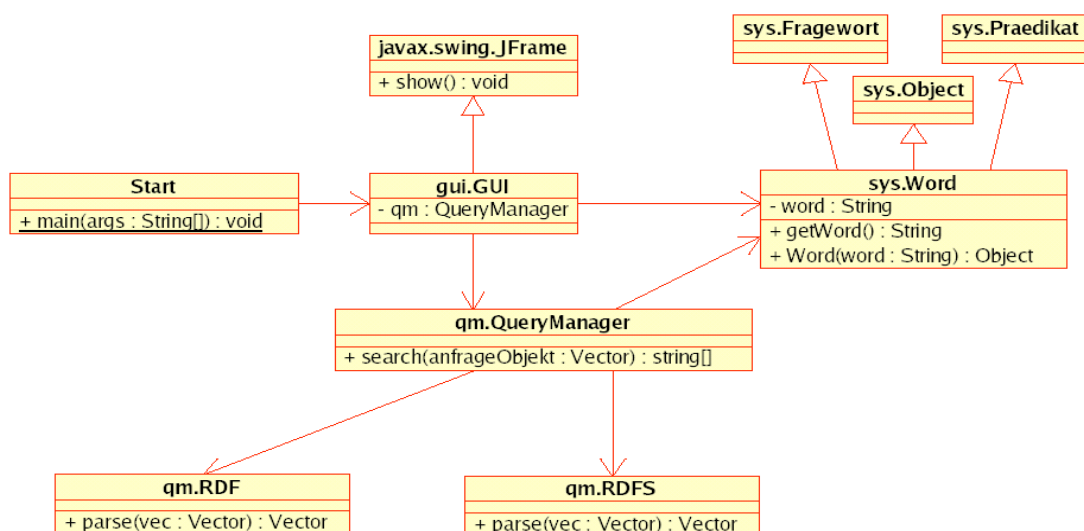
Vizitator	patina		afara		Brasov	URL
Vizitator	patina		in sala		Cluj	URL

If action_area is not found in the DB select childs of "Transilvania" from the result set.

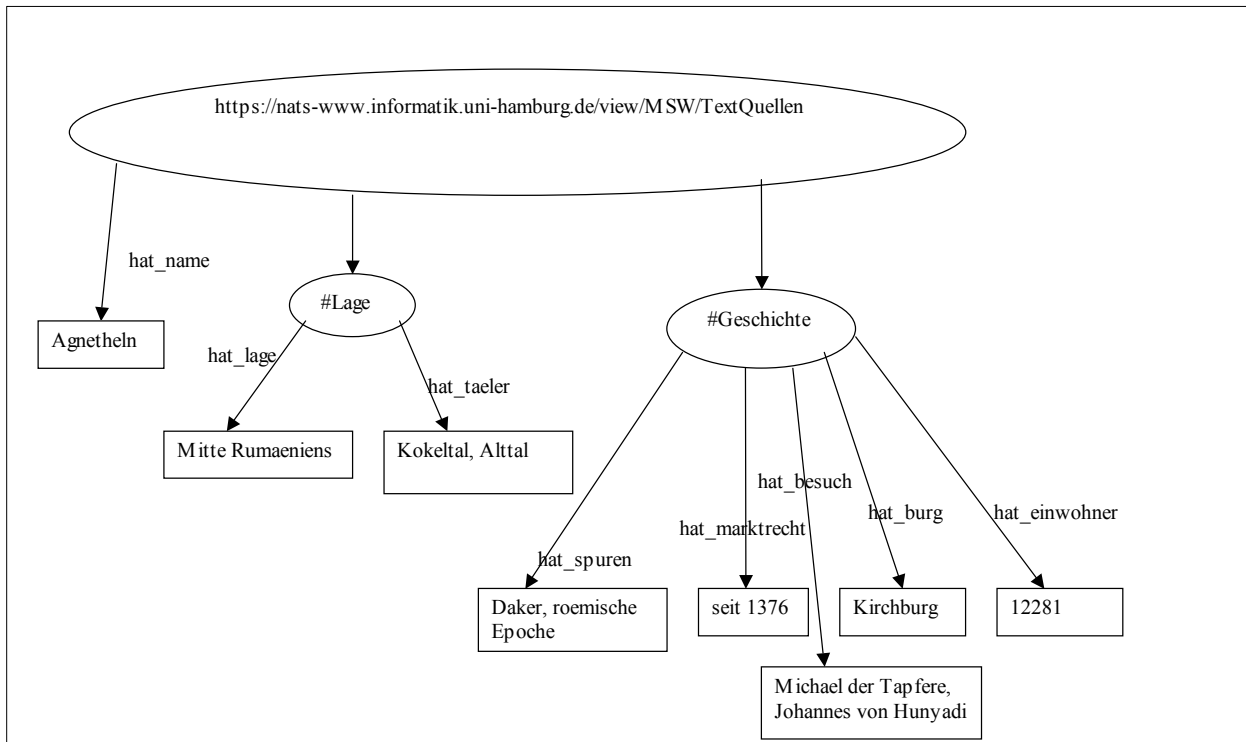
Query Manager (Ontology (RDFS) and RDF Data)-1-



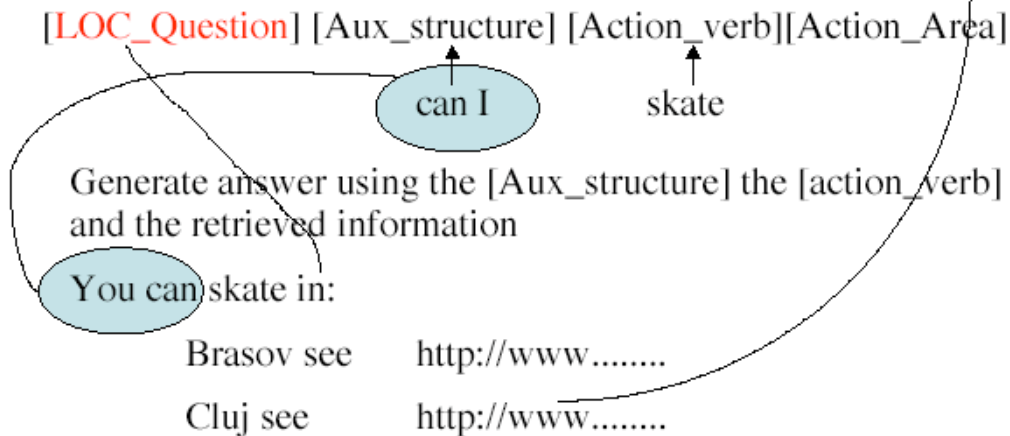
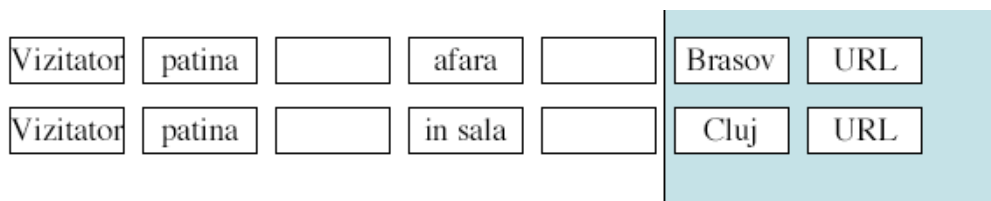
Query Manager (Ontology (RDFS) and RDF Data)-2-



RDF -Representation



Generation of the Answer -1-



Generation of the Answer -2-

No Result?

- *"We do not have any specific answer for your request" or*
- Give alternatives with the specified action_verb
 - *"You can only fly to Cluj, Sibiu and Timisoara"*
- Give URLs of all cities \subset Transylvania.
- Ask about near alternatives:
 - *"Would it fit to your plans to do skiing in the mountains?"*
- or combinations of them

Conclusions and further work

- For the moment it is a didactical system
- The approach was tested on a corpus of several hundred questions and a collection of about hundred texts (German-English)
- Further Plans:
 - to encode the ontology in OWL
 - Management of multilingual mapping problems
 - Extend the coverage of the system

