STU FIIT

SLOVAK UNIVERSITY OF TECHNOLOGY IN BRATISLAVA
FACULTY OF INFORMATICS AND INFORMATION TECHNOLOGIES

keystone



Using Natural Language to Search Linked Data

Ing. Peter Macko

Supervisor: Assoc. Prof. Viera Rozinajova

2nd International KEYSTONE Conference

Our Problem

- Asking question using natural language
- Utilize the semantic web
  - SPARQL – to complicated
    - Structure has to be know by user
    - Knowledge about syntax
  - Weak support of search engines



Examples

- Which German cities are birth places of physicists with Nobel Price?

```
SELECT
  ?city
WHERE {
  ?city     dbo:country      dbr:Germany.
  ?person   dbo:birthPlace   ?city.
  ?person   dbo:type         yago:Physicist110428004.
  ?person   dbo:award        ?award.
  ?award    dbo:subject      dbr:Nobel_Price
}
```

Nobel Prize - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/**Nobel_Prize**
The **Nobel Prize** is a set of annual international awards bestowed in a number of categories by ... The **prize** ceremonies take **place** annually in Stockholm, Sweden, except for the peace **prize** which is held in ... The Royal Swedish Academy of Sciences awards the **Nobel Prize** in **Physics**, the **Nobel Prize** in Chemistry, and the ...

Science and technology in Germany - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Science_and_technology_in_**Germany**
Germany's achievements in science and technology have been very significant and research and development efforts form an integral part of the country's economy. **Germany** has been the home of some of the most prominent researchers in various scientific disciplines, notably **physics**, ... Before World War II, **Germany** had generated more **Nobel laureates** in ...

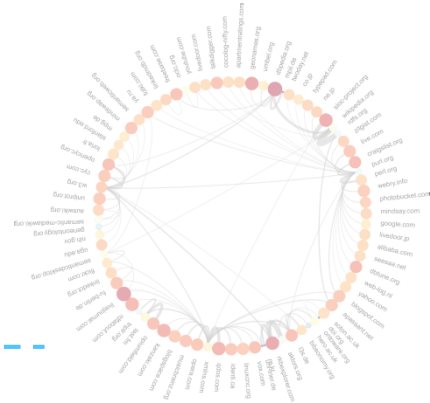
List of Nobel laureates by country - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/List_of_**Nobel_laureates**_by_country
This is a list of **Nobel Prize laureates** by country. Listings for Economics refer to the related ... For the same award, two **birthplace** related listings occur when the **place** of **birth** is currently in a different country than at the time of **birth**. ..... born in then Free **City** of Danzig, now Poland, Literature, 1999; Horst L. Störmer, **Physics**, ...

Nobel Laureates and Country of Birth
https://www.**nobelprize**.org/**nobel_prizes**/lists/**countries**.html
Austria, **Germany**, Poland and Russia are other examples of **countries** where the borders have ... Interactive world map showing **Nobel Laureates** and country of **birth** ... The Nobel Prize in **Physics** 1915 ..... Free **City** of Danzig (now Poland) (1).

The Nobel Prize - Sweden.se
sweden.se › Society
Jan 7, 2016 - The **Nobel Prize** is considered the world's most prestigious **award**. ... the **Laureates** in **Physics**, Chemistry, Physiology or Medicine, and Literature ... The four institutions in Sweden and Norway (the two **countries** were united ... **Nobel** Peace **Prize** was awarded to Carl von Ossietzky, a **German** writer who ...

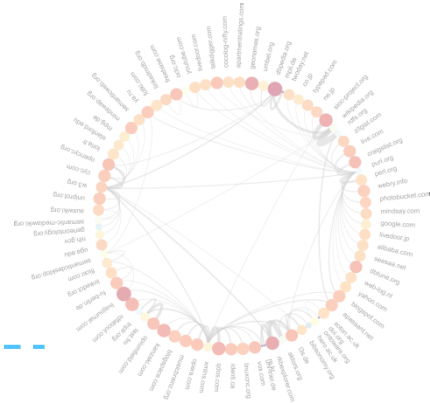
State of the Art

- Firstly defined on SQL
- Controlled natural language
  - Templates
  - Menu
- Suggestions
- Sentence analyzers
- Preprocessing
  - Word databases



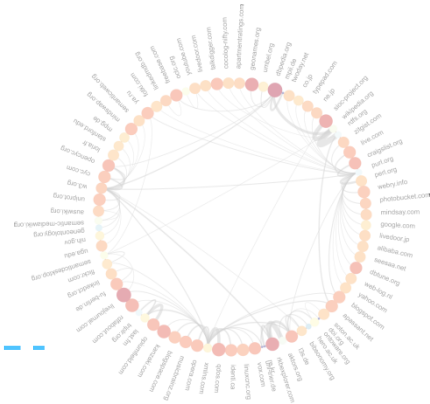
Main Problems

- Vocabulary problem
  - Datasets doesn‘t contains alternative phrases
- Homonyms
  - Jaguar vs. jaguar
- Sentence structure



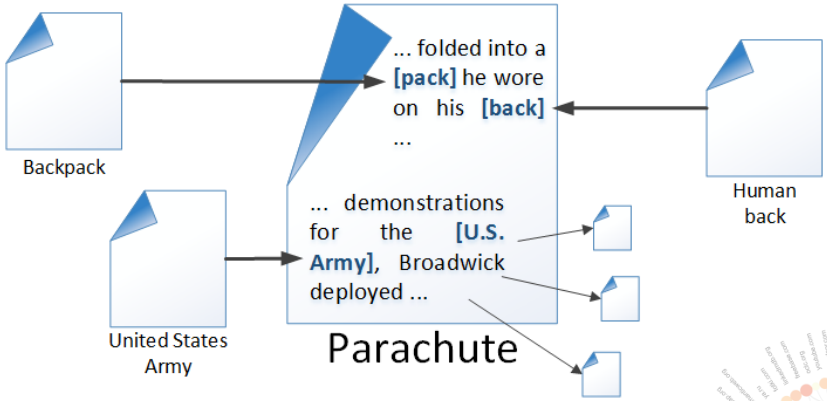
Our Method

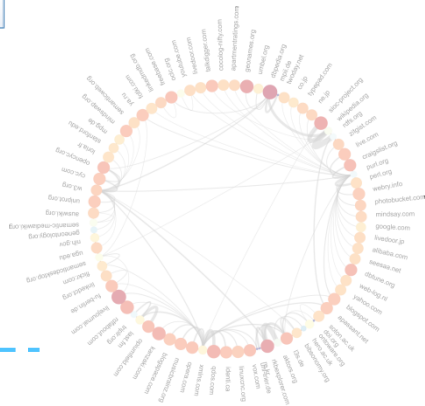
- Preprocessing
  - Analysis of dataset
  - Collecting alternative phrases in lexicons with its weights
- Query time
  - Sentence analysis
  - Lookup for the path in dataset



Preprocessing

- Dataset analysis
- Alternative phrases lookup
  - Wikipedia
  - Word database
    - Synonyms
    - Multiword phrases

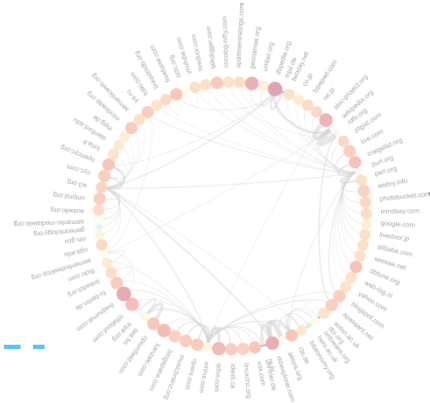




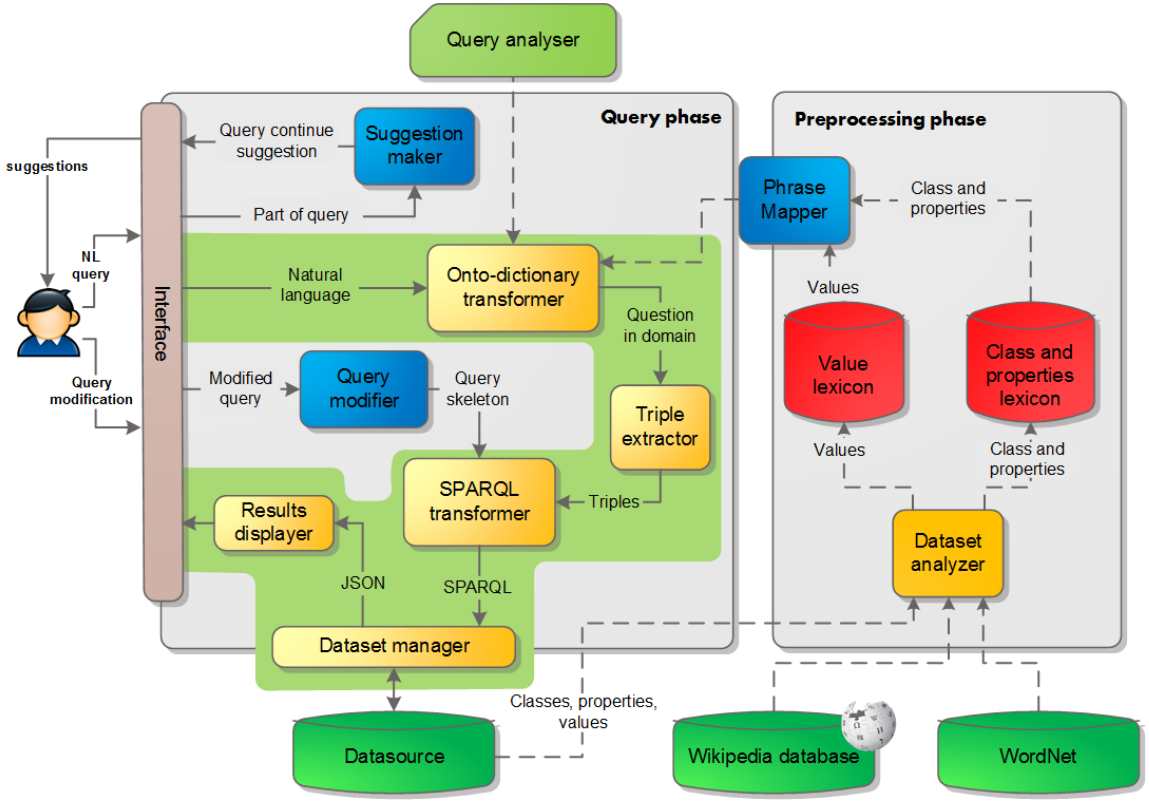
Lexicon weighting

- Many possible translations
- All alternative phrases get its weight:
  - $W_{descriptor} = w_{descriptor\ source} \times (upperclass_const)^{nesting\ degree}$

	Picture	Photo
Picture	1	0.3
Image	0.7	0.7
Photo	0.6	1



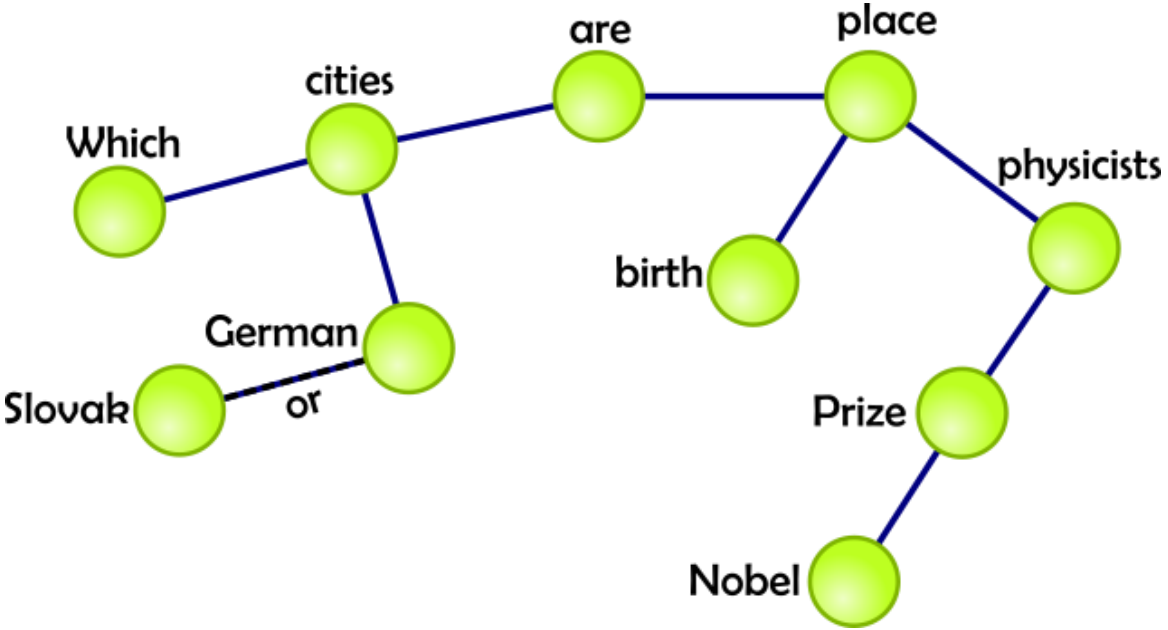
Method schema

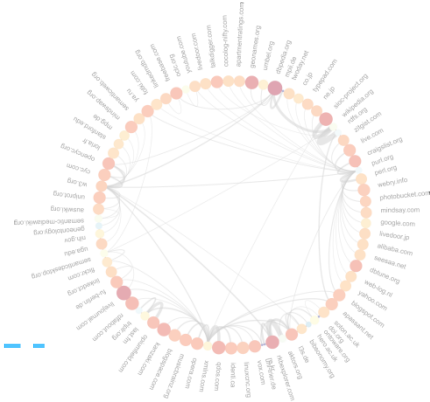




Example

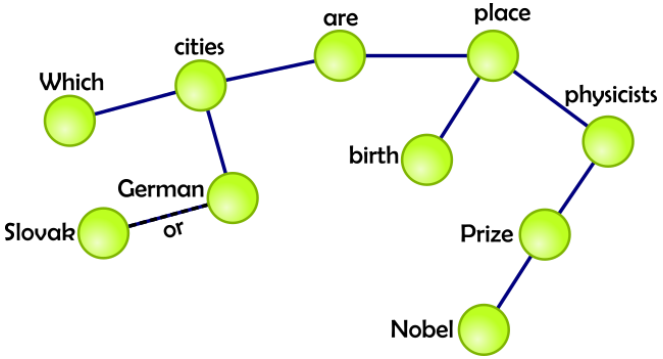
Which German or Slovak cities are birth places of physicists with Nobel Prize?

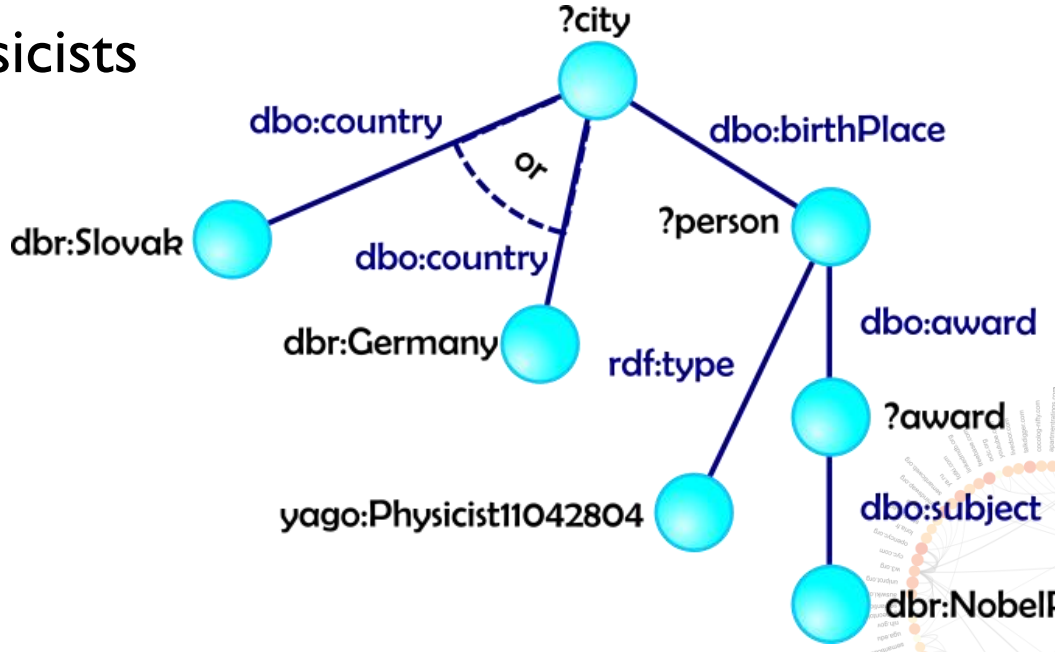




Example

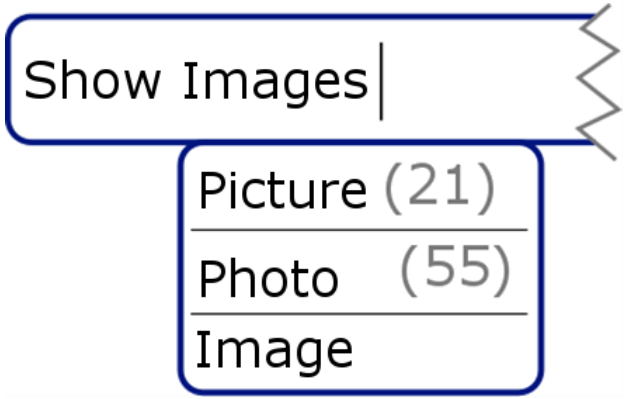
Which German or Slovak cities are birth places of physicists with Nobel Prize?



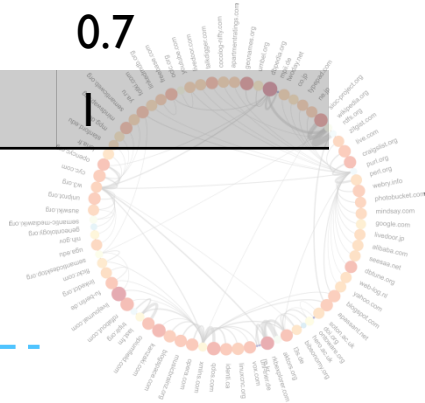


Suggestions

- Helps user to create query
- Ordering prefers namings fom dataset



	Picture	Photo
Picture	1	0.3
Image	0.7	0.7
Photo	0.6	1



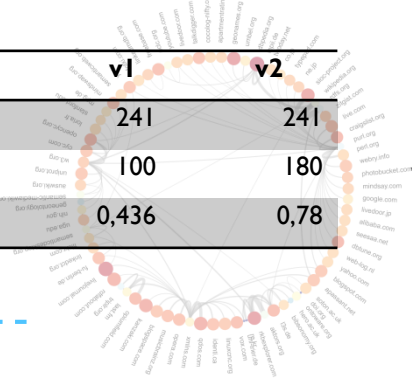
Results

- Experiments with preprocessing stage
- Experiments on overall method
  - Usage of first 50 queries form NL-SPARQL database

Match	Ordrings (%)
exact	90
1 mistake	6
3 mistakes	4

	v1	v2
Questions	50	50
Queries	19	39
Satisficatory	14	36

	v1	v2
Triples	241	241
Right	100	180
F1 score	0,436	0,78



Future Work

- Focus on group operation
  - Highest peek
  - Longest river
- Comparison with other methods using challenge

Conclusion

- Natural language as a query language
- Dataset preprocessing
  - Data collecting
  - Weight calculation
- Sentence analysis

