

LingNLQ: Natural Language Query for linguistics

M2 TAL Software Project

Karolin Boczoń
Roham ROSHANFEKR

by

Dimitra NIAOURI
Muhammad SHAHZAIB

Linguistics domain issues

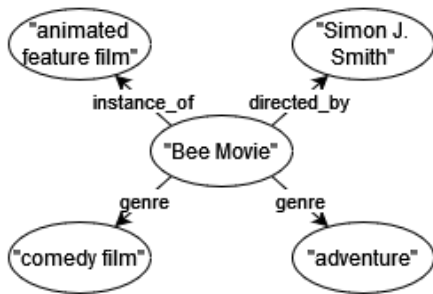
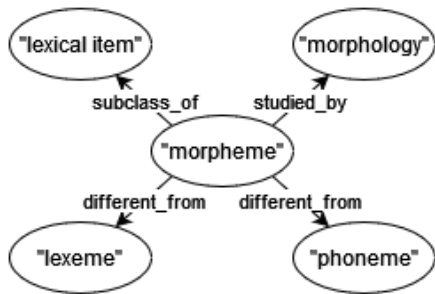


Figure: Linguistics concepts vs real-life entities

Linguistics domain issues

The domain (linguistics) and the resource (knowledge graph) we chose aren't compatible.

Possible new project ideas:

- ▶ Closed-domain Information Retrieval-based Question Answering
- ▶ **Open-domain Knowledge-based Question Answering**

TriviaNLQ: Natural Language Query for trivial questions

M2 TAL Software Project

Karolin Boczoń
Roham ROSHANFEKR

by

Dimitra NIAOURI
Muhammad SHAHZAIB

Why trivia?

- ▶ trivia is little-known and rather useless facts
- ▶ mostly used in trivia quizzes
 - ▶ sets of questions available
 - ▶ a good trivia question is unambiguous and objective
- ▶ suitable for knowledge-based approach

TriviaQA

- ▶ reading comprehension dataset
- ▶ includes 95K question-answer pairs authored by trivia enthusiasts
- ▶ comes with an evaluation script

Question	Which musical instrument does Kermit the Frog play?
Answer	banjo

Table: Example pair

Translating to query

Which *musical instrument* does **Kermit the Frog** play?

```
1 SELECT ?item ?itemLabel WHERE {  
2 # Kermit      instrument  
3   wd:Q1107971 wdt:P1303 ?item .  
4   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }  
5 }
```



1 result in 311 ms

</> Code

Download ▾

Link ▾

item

itemLabel

wd:Q258896

banjo

Text to QDRM to SPARQL

- ▶ the Model was Originally SQL based
- ▶ steps for model evaluation and testing
- ▶ SPARQLing Package
<https://github.com/yandex-research/sparqling-queries>

SPARQLing Queries

- ▶ approach of QDMR2SPARQL
- ▶ steps for model evaluation
- ▶ understanding scripts

What's next

- ▶ adapting Text2QDRM to *TriviaQA* questions
- ▶ adapting QDRM2SPARQL to *WikiData* schema
- ▶ model evaluation
- ▶ building the website

Timeline

Week No.	Milestone
46	Run sparqling-queries
55	First draft of the report
02	Finished all the experiments
05	Finished report
05	Deployed website

Subject to change.

Thank you!