

# LingNLQ: Natural Language Query for linguistics

M2 TAL Software Project

by

Karolin BOCZOŃ

Dimitra NIAOURI

Roham ROSHANFEKR

Muhammad SHAHZAIB

# Idea

Question answering system related to the domain of linguistics.

- ▶ queries in natural language
- ▶ knowledge base all about language and linguistics
- ▶ using Text2SPARQL model tuned with terminology

*What language has /θ/ sound?*

# Market research



how many cases does greek have?



NATURAL LANGUAGE



MATH INPUT



EXTENDED KEYBOARD



EXAMPLES



UPLOAD



RANDOM

Input interpretation

Greek

case

Definitions »

Result

yes (bicameral script)

# Market research

how many cases does greek have



All

News

Images

Videos

Books

More

Tools

About 199,000,000 results (0.53 seconds)

## five CASES

29. There are **five** CASES in Greek, the nominative, genitive, dative, accusative, and vocative. In English, readers rely on the order in which words appear in a sentence to indicate the grammatical function of each word. In Ancient Greek, their case tells the reader the grammatical function of each word in the sentence.

<https://daedalus.umkc.edu> › FirstGreekBook › JWW\_FGB3 ›

[Lesson 3. Nouns.—Introductory](#)

# Team

- ▶ Karolin (CS) - project managing
- ▶ Dimitra (ling) - expertise in linguistics
- ▶ Roham (telecom) - software integration
- ▶ Shahzaib (CS) - web development

Useful skills acquired during master's:

- ▶ machine learning
- ▶ web semantics
- ▶ scraping

# Technical stuff

# Technology stack

- ▶ backend - Python
- ▶ frontend - JavaScript
- ▶ task management - GitHub Projects
- ▶ query language - SPARQL
- ▶ storage - Blazegraph



# Application model

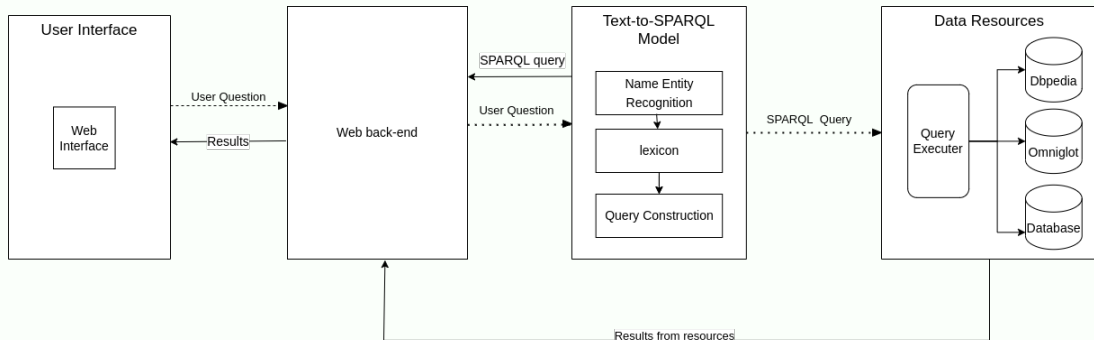


Figure: Structure of the Web Application

# User interface



A screenshot of a web browser displaying the LingNLQ interface. The browser's address bar shows 'lingnlq.com'. The page features the title 'LingNLQ' in a large, black, sans-serif font. Below the title is a light gray input field with a question mark icon on the left. Underneath the input field is an orange button with the text 'ask!'. Below the button are two stacked light gray rectangular boxes; the top one contains the text '[interpretation of query]' and the bottom one contains '[answer]'. The entire interface is framed by an orange border.

# Data resources

- ▶ RDF schema:
  - ▶ Wikidata (<https://www.wikidata.org/>)
  - ▶ DBpedia (<https://www.dbpedia.org/>)
- ▶ Linguistics & language data:
  - ▶ Omniglot (<https://omniglot.com/>)
    - maintained by an expert, free for personal use
  - ▶ World Atlas of Languages (<https://wals.info/>)
    - large database from reference material, CC BY 4.0
  - ▶ Ethnologue (<https://www.ethnologue.com/>)
    - lots of data however not open source

# Text2SPARQL

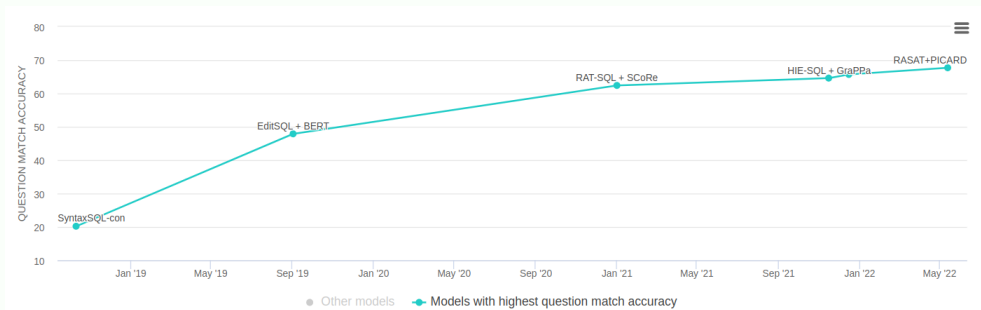


Figure: Available SOTA Models

The best of the best models has at most 70% of question matching accuracy.

# Thank you!