TEXT2SPARQL

TREx (Text-to-RDF Executor)

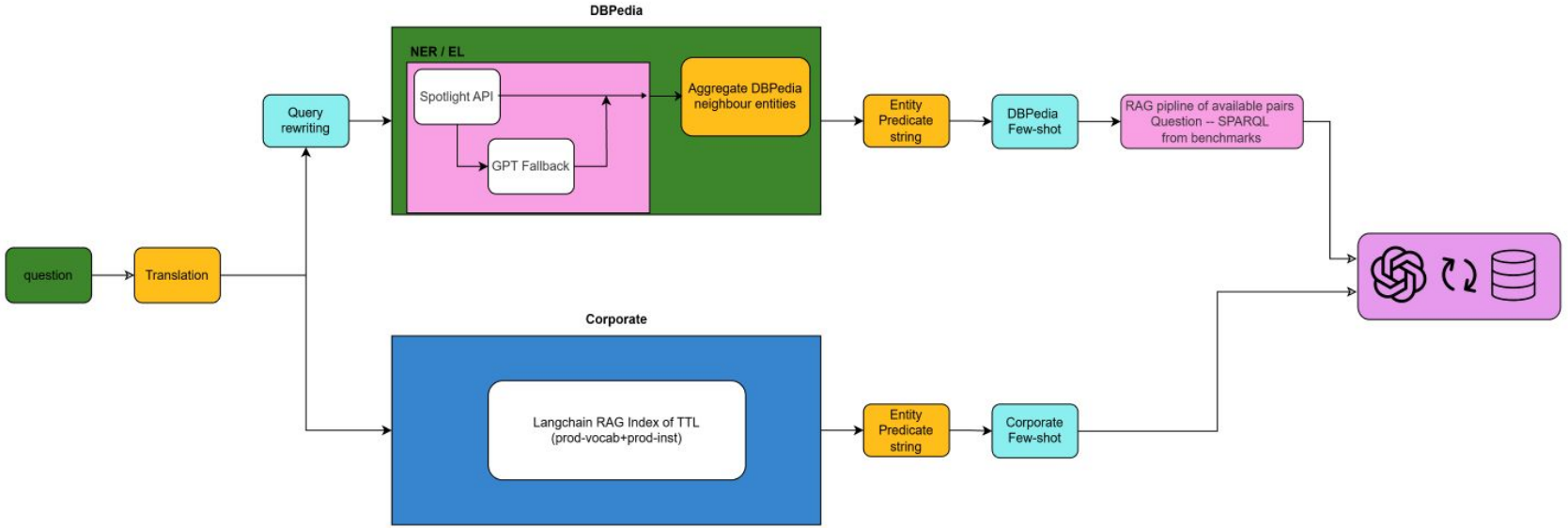
Somov Oleg, AIRI
Berezin Daniil, MIPT
Avdeev Roman, MIPT

Intro

What are the problems in the text2sparql task

- Multilingual complexity (GPT translation)
- Entity linking accuracy (Query rewriting and Spotlight)
- Knowledge Graph ontology alignment ("Nearest neighbors" and RAG)
- SPARQL query generation (Adaptive prompt)

Architecture

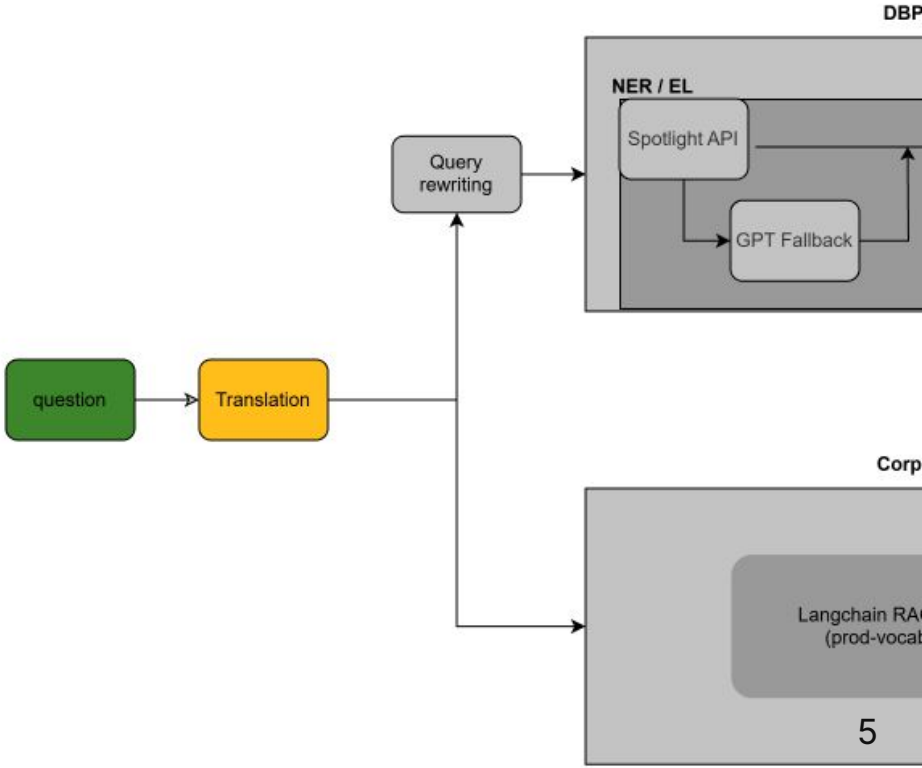


Baseline

- As a simple baseline we have used **ChatGPT** with few-shot examples
- The baseline execution match accuracy – **17%**
- For evaluation of DBPedia we have used a subset from LcQuad2.0 and QALD-10 (multilingual)
- Execution on DBPedia via API
- Execution on Corporate via OpenLink Virtuoso graph database

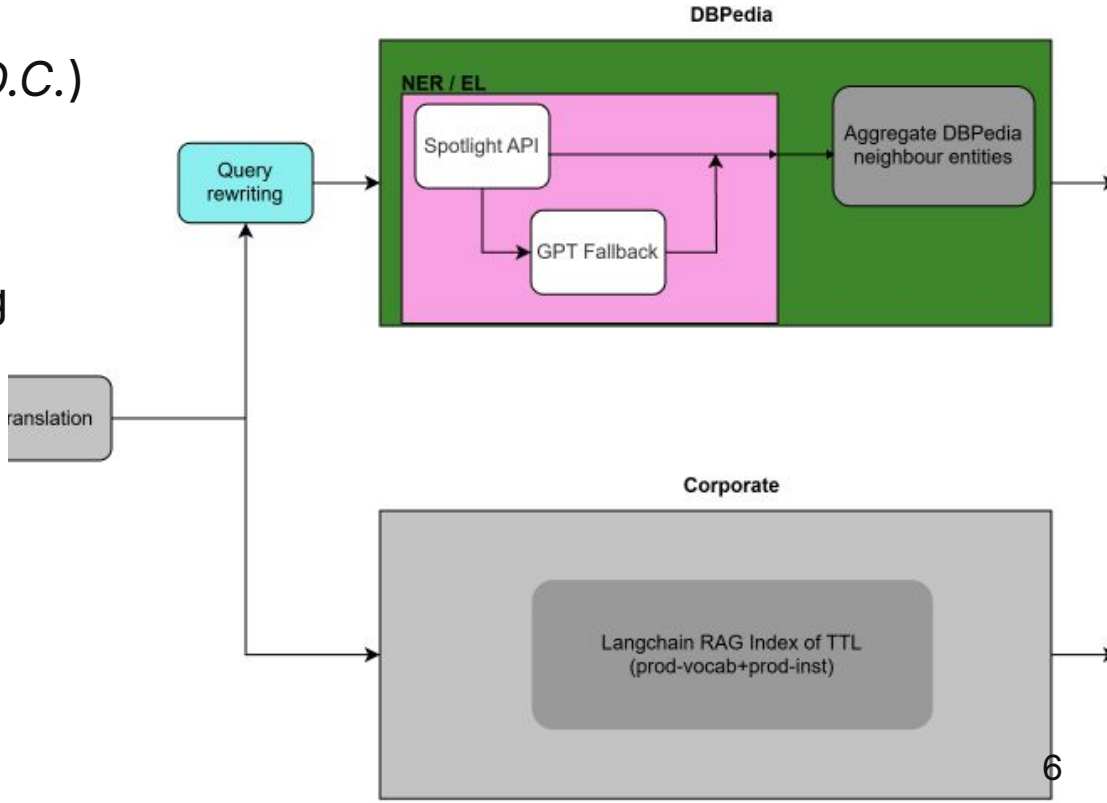
Multilinguality problem

- Problem - The model does not understand entities well in other languages.
- Solution - add English translation via GPT.
- Uplift - +**7**% execution match



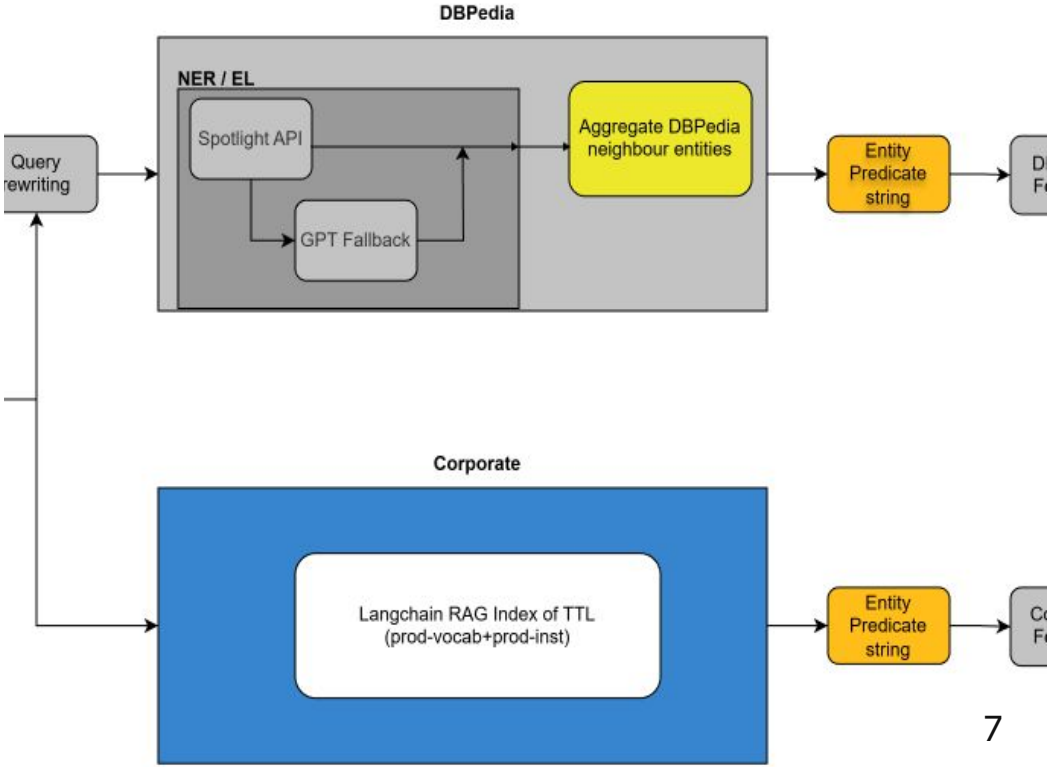
Entity linking problem

- Solution - Query rewriting (*Washington* vs *Washington D.C.*)
- DBPedia Spotlight for Entity Recognition and Entity Linking
- Uplift - +**6**% execution match



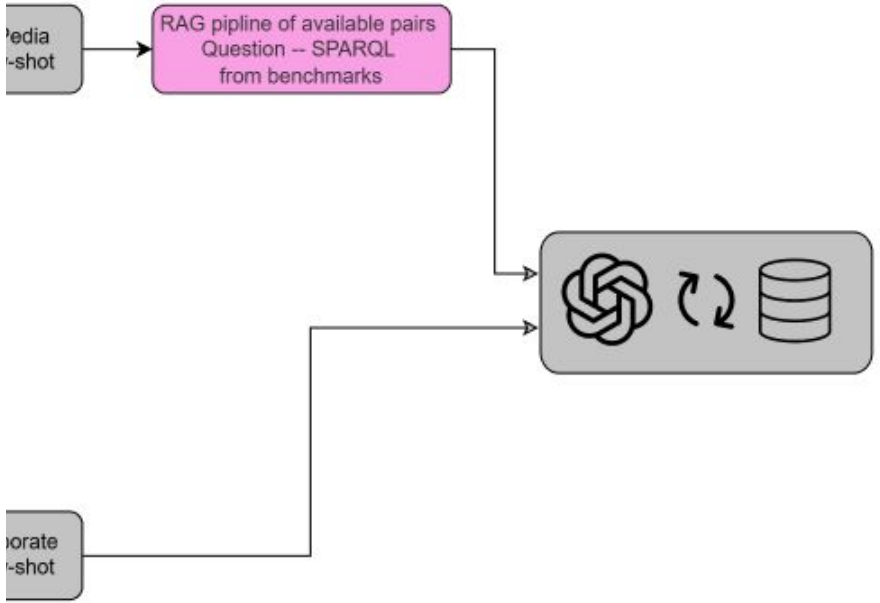
LLM does not know the ontology of underlying graphs

- To generate a high-quality LLM response, you need an understanding of the knowledge graph.
- To do this, we search for "nearest neighbors" in DBpedia.
- For the corporate graph, we create a RAG based on TTL files.
- Uplift - +**12**% execution match



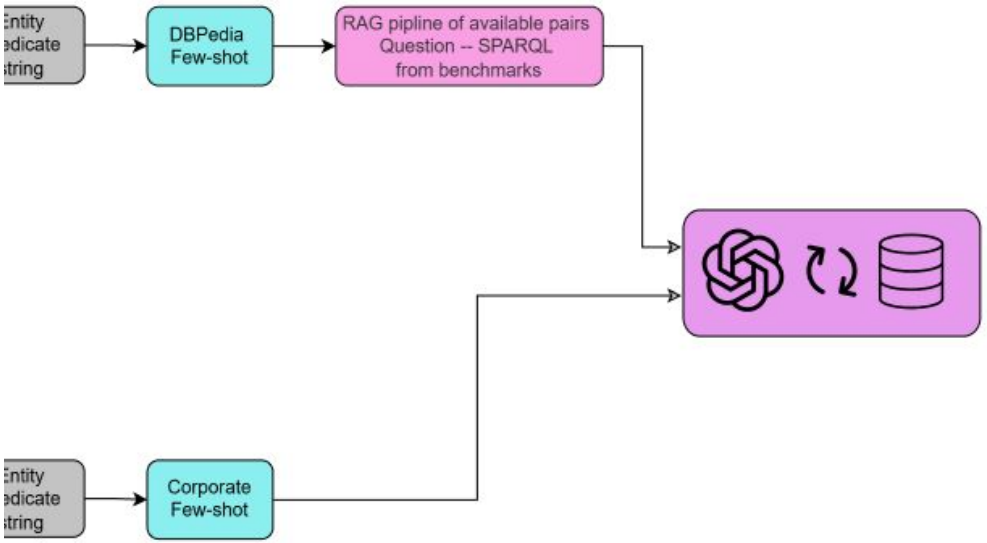
Adaptive prompt

- Adaptive prompt was used to improve generation.
- For this, RAG was used on several popular DBpedia and Wikidata datasets (quald9b, lcquad, and others).
- Uplift - +**6**% execution match

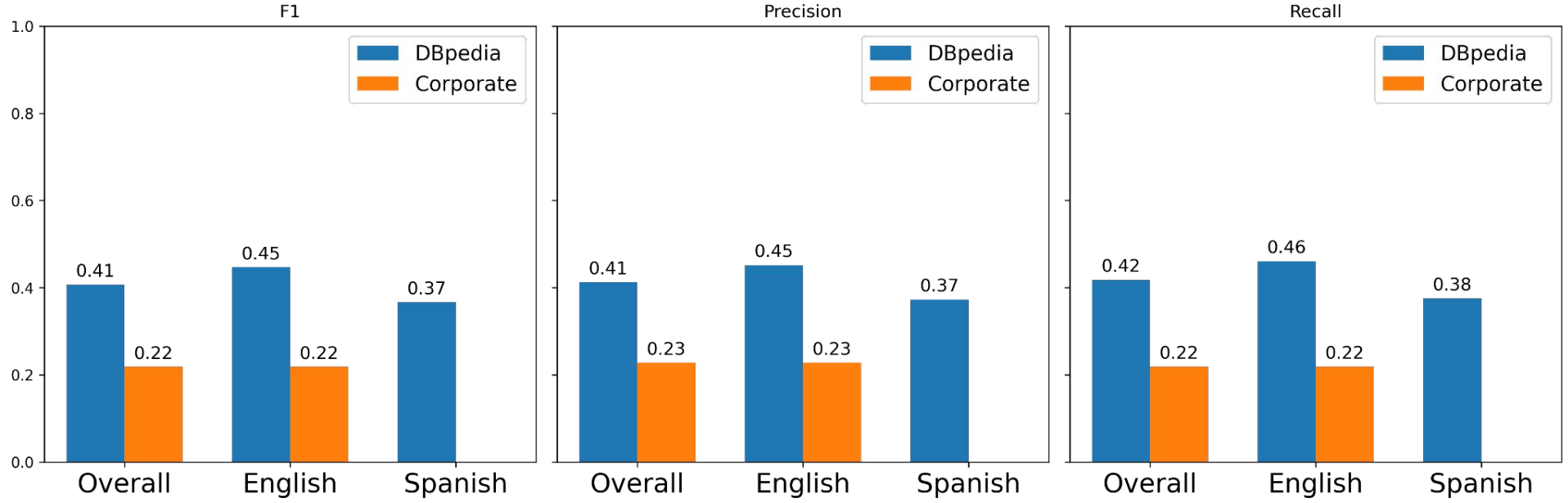


How do query generation

- We augment the input to ChatGPT with relevant question-query pairs from RAG
- Index consists from **LCQuad** and **QALD** question-query pairs
- With the relevant entities, predicates, reasoning few-shots and mined via RAG exemplars along with the input question we generate a query
- We do execution guided generation (5 tries) to generate an executable SPARQL



Final results



Team

Somov Oleg
somov@airi.net

Berezin Daniil
berezin.da@phystech.edu

Roman Avdeev
roma.avdeyev@gmail.com

AIRI MIPT MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY

Links

Github