Aiteam
NATIONAL & KAPODISTRIAN
UNIVERSITY OF ATHENS

Laboratoire
Hubert Curien
UMR • CNRS • 5516 • Saint-Étienne

The Question Answering System GeoQA2

Dharmen Punjani, Sergios - Anestis Kefalidis, Konstantinos Plas,
Eleni Tsalapati, Manolis Koubarakis and Pierre Maret

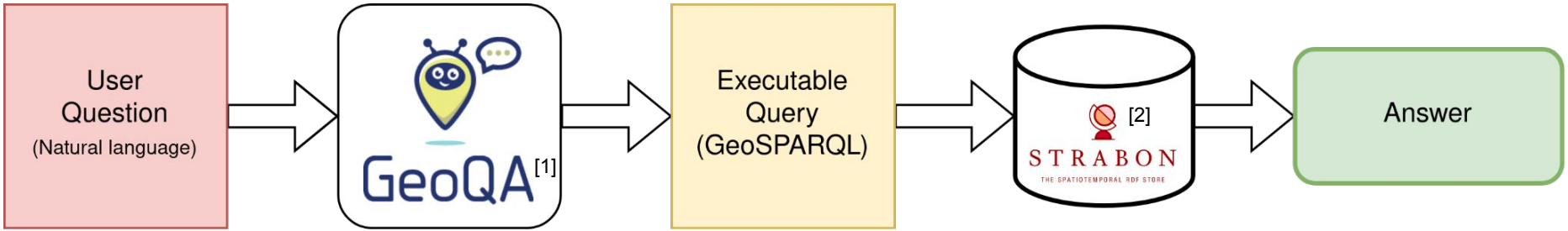
Outline

- Previous work on Geospatial Question Answering
- GeoQA2
- Evaluation
- Future work

Previous work on Geospatial Question Answering

- Hybrid geo-spatial query methods on the semantic web with a spatially-enhanced index of DBpedia.
  - Younis et al., 2012
- **Template-based question answering over linked geospatial data.**
  - Punjani et al., 2018 and 2020
- Neural factoid geospatial question answering.
  - Li et al., 2021
- Qualitative spatial reasoning over questions.
  - Beydokhti et al., 2022
- **Translating place-related questions to GeoSPARQL queries.**
  - Hamzei et al., 2022

Overview of the Question Answering Process with GeoQA



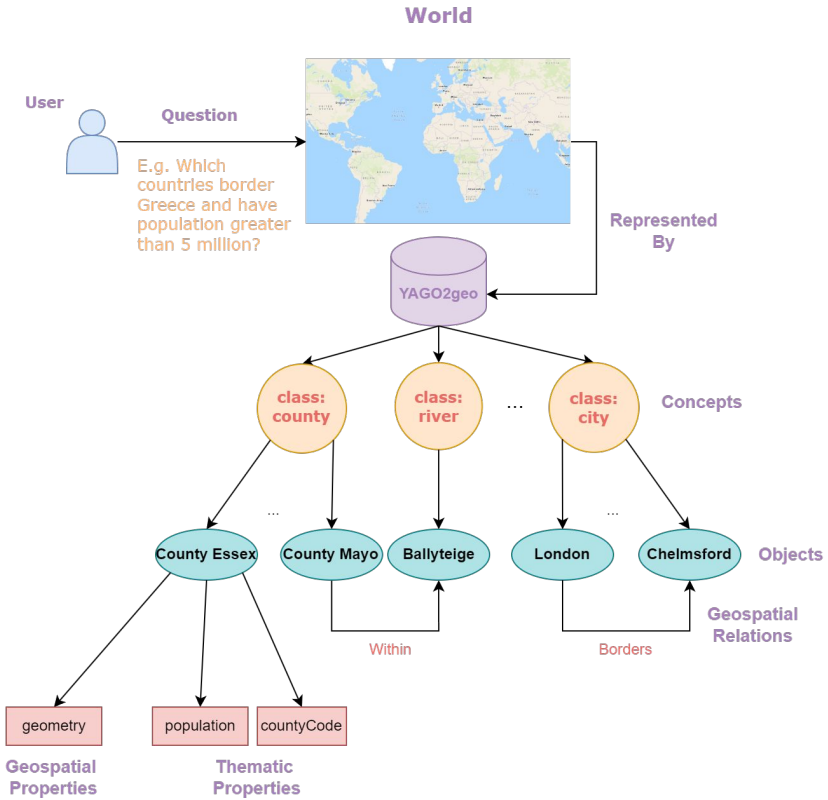
[1] Dharmen P., Kefalidis S.-A., Plas K., Tsalapati E., Koubarakis M., Maret P. *The Question Answering System GeoQA2.* Geospatial Knowledge Graphs and GeoAI: Methods, Models, and Resources - 2023
[2] Kyzirakos K., Karpathiotakis M., Koubarakis M. *Strabon: A Semantic Geospatial DBMS.* International Semantic Web Conference - 2012

Improvements in GeoQA2

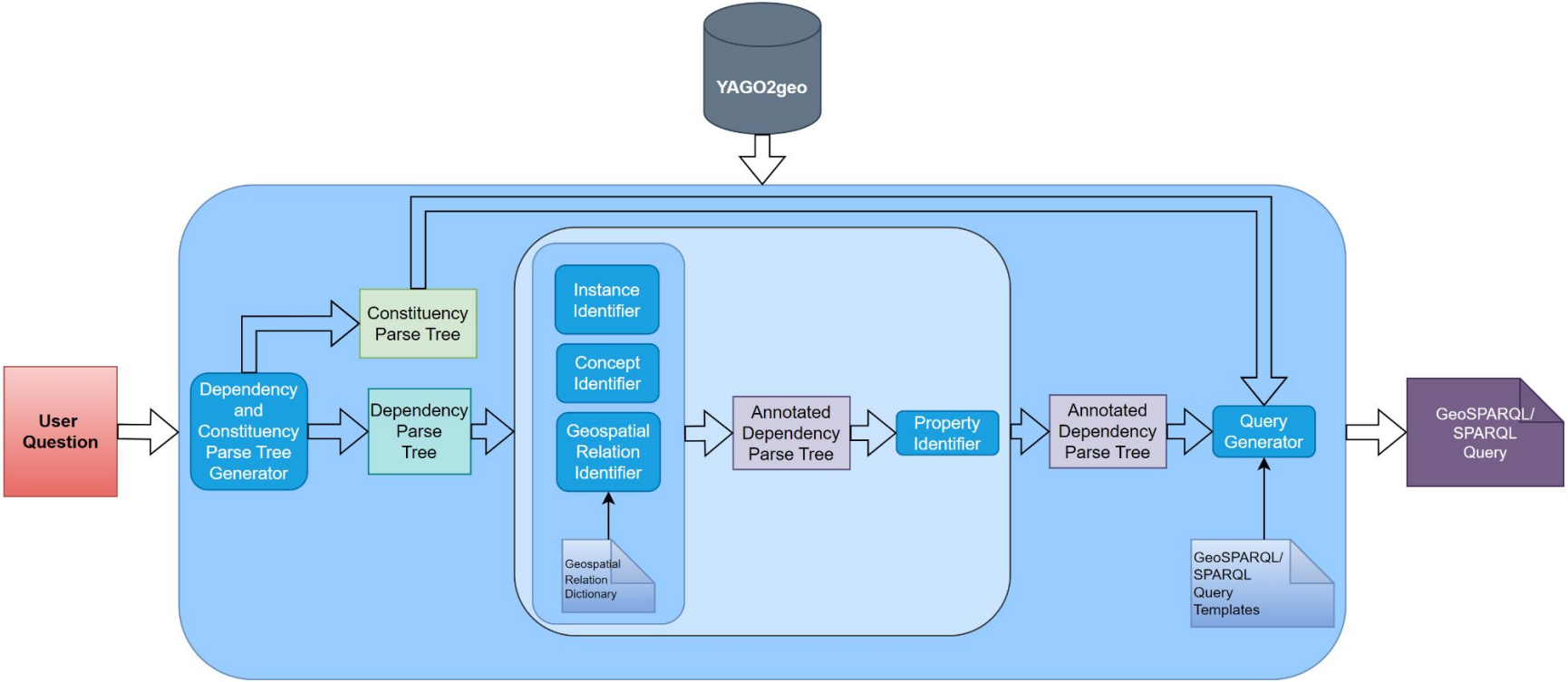
- Different target knowledge graph:
  - GeoQA:
    - **DBpedia [1]**
    - **OpenStreetMap [2] (**UK and Ireland)
    - **Global Administrative Areas [3]** (UK and Ireland)
  - GeoQA2:
    - **YAGO2 [4]**
    - **YAGO2geo [5]** (UK, Ireland, USA and Greece)
- More supported question types:
  - **Quantities**: How many lakes are there in Greece?
  - **Aggregates**: What is the total area of lakes in Monaghan?
  - **Comparatives**: Is Lake Baikal bigger than the Ioannina lake?
- Optimized modules.

[1] https://www.dbpedia.org/
[2] https://www.openstreetmap.org/
[3] https://gadm.org/
[4] https://yago-knowledge.org/
[5] https://yago2geo.di.uoa.gr/

Knowledge Representation in GeoQA2

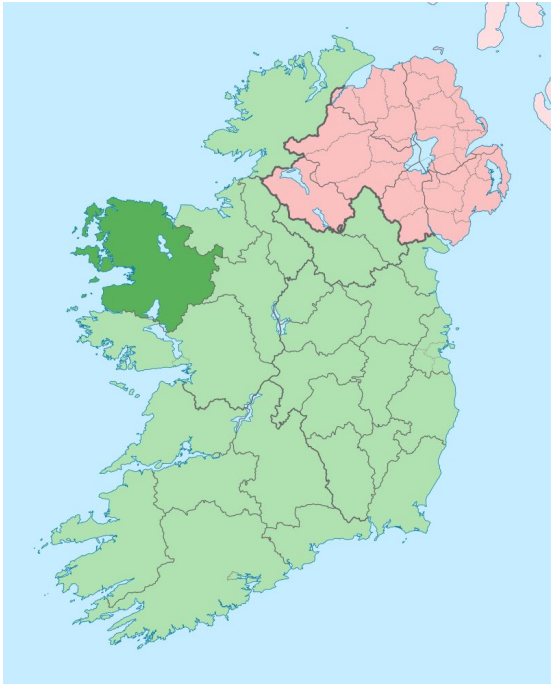


The GeoQA2 Pipeline: Overview



The GeoQA2 Pipeline: Running Example

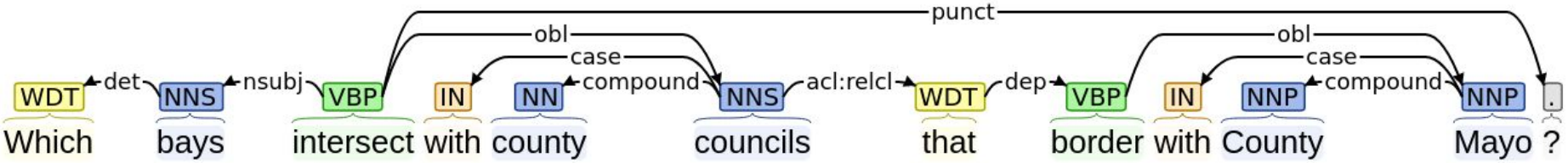
- Which bays intersect with county councils that border with County Mayo?



The GeoQA2 Pipeline: Dependency and Constituency Parse Tree Generator (1)

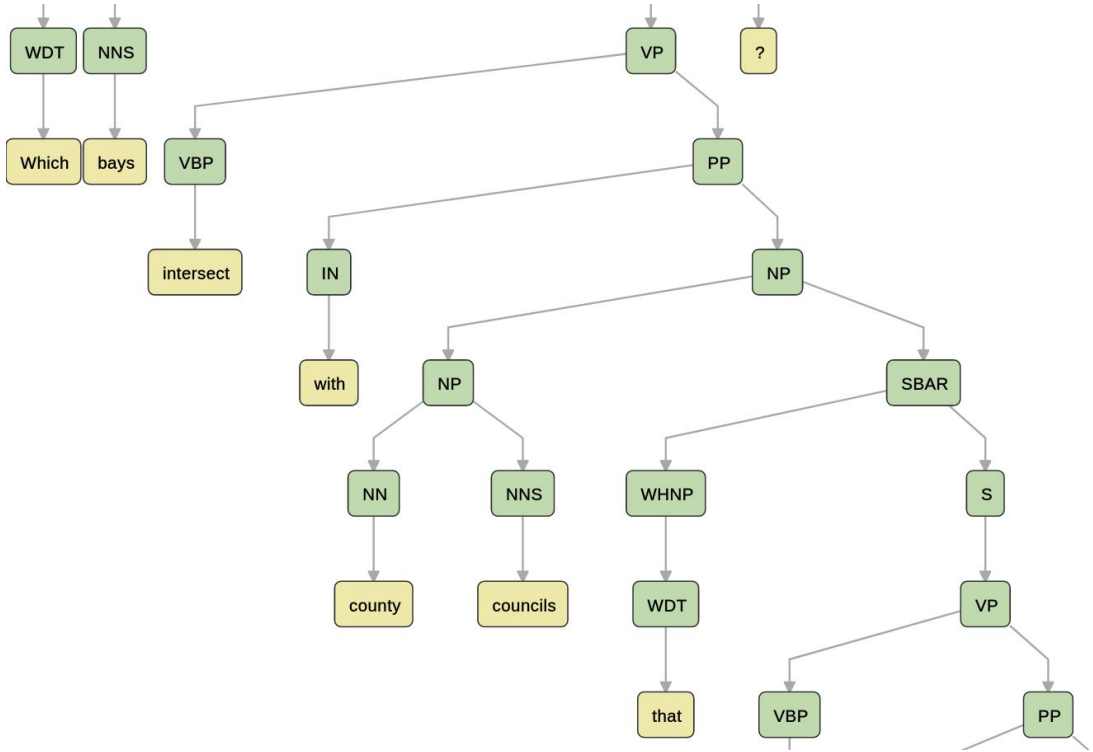
This component carries out:

- **part-of-speech tagging**
- **dependency parse tree** generation (annotated by pipeline components)
- **constituency parse tree** generation (used by the query generator)



Dependency Parse Tree

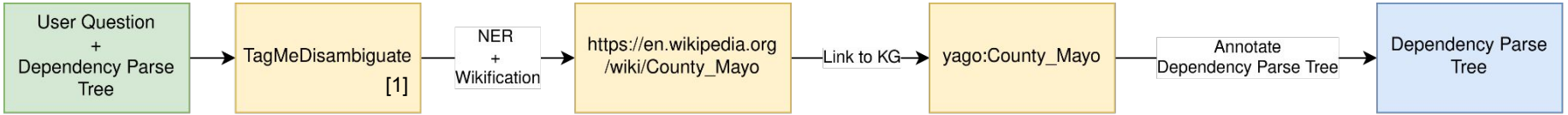
The GeoQA2 Pipeline: Dependency and Constituency Parse Tree Generator (2)



Partial Constituency Parse Tree

The GeoQA2 Pipeline: Instance Identifier

- Identifying **geographic features** (instances), e.g. County Mayo, and linking to the target KG.
- How it works:



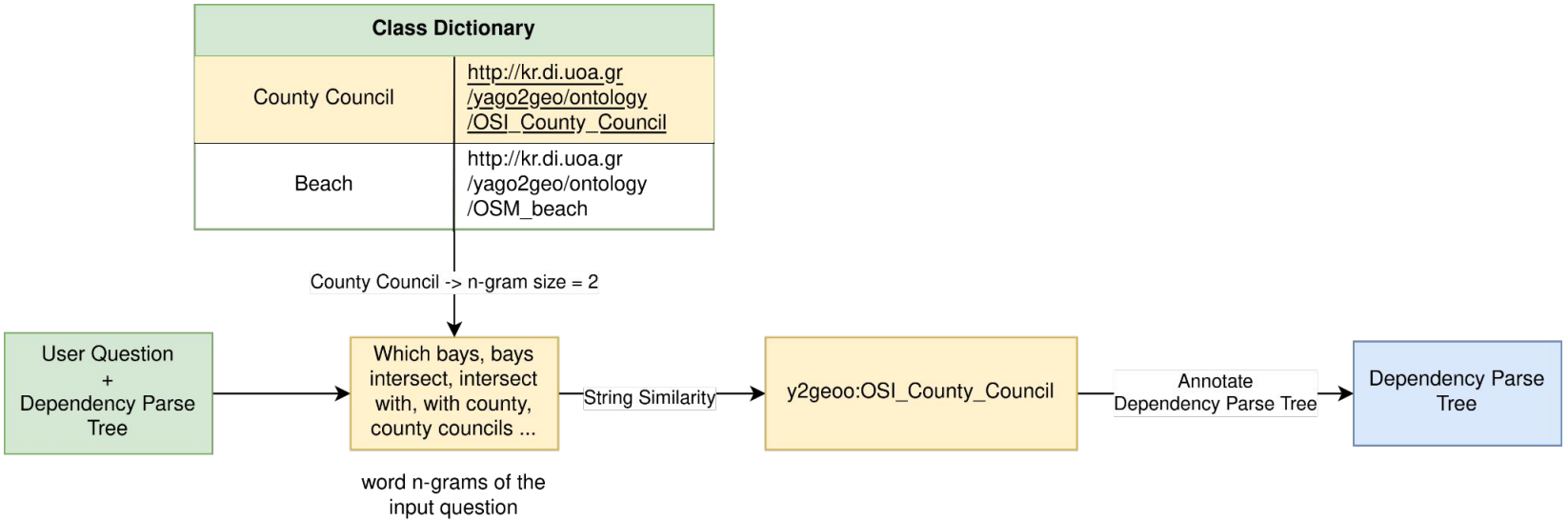
[1] Ferragina P., Scaiella U. *TAGME: on-the-fly annotation of short text fragments (by wikipedia entities).* CIKM - 2010

The GeoQA2 Pipeline: Running Example

- Which bays intersect with county councils that border with County Mayo?
  - Instances: yago:County_Mayo

The GeoQA2 Pipeline: Concept Identifier

- Identifying **types of geographic features** (concepts), e.g. forest, and mapping to the ontology of the KG.
- How it works:

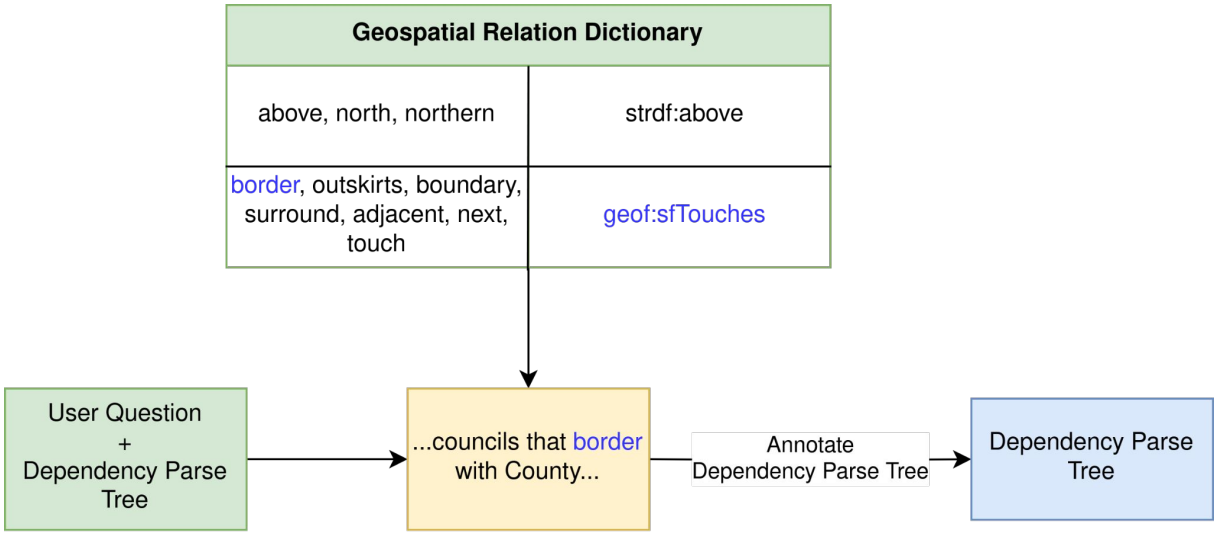


The GeoQA2 Pipeline: Running Example

- Which bays intersect with county councils that border with County Mayo?
  - Instances: yago:County_Mayo
  - Concepts: y2geoo:OSI_County_Council, y2geoo:OSM_bay

The GeoQA2 Pipeline: Geospatial Relation Identifier

- Identifying **geospatial relations**, e.g. borders, and mapping to GeoSPARQL/stSPARQL functions.
- How it works:

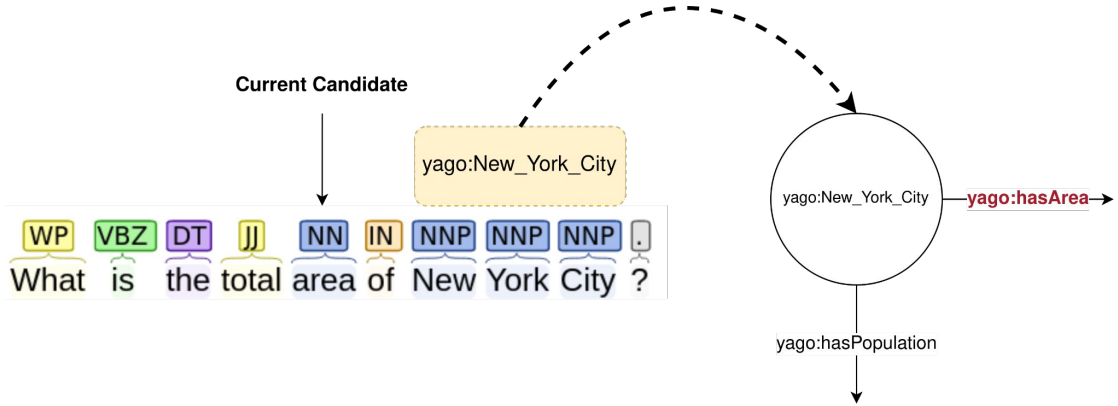


The GeoQA2 Pipeline: Running Example

- Which bays intersect with county councils that border with County Mayo?
  - Instances: yago:County_Mayo
  - Concepts: y2geoo:OSI_County_Council, y2geoo:OSM_bay
  - Relations: geof:sfIntersects, geof:sfTouches

The GeoQA2 Pipeline: Property Identifier

- Identifying **attributes of types of geographic features** and **attributes of geographic features** (properties), e.g. population, and mapping them to the target KG.
- How it works:
  - Identify potential properties: POS tags {NN, JJ, NNP, NP}.
  - Match with KG properties (1-hop relations) of the **Concepts** and **Instances** identified previously.

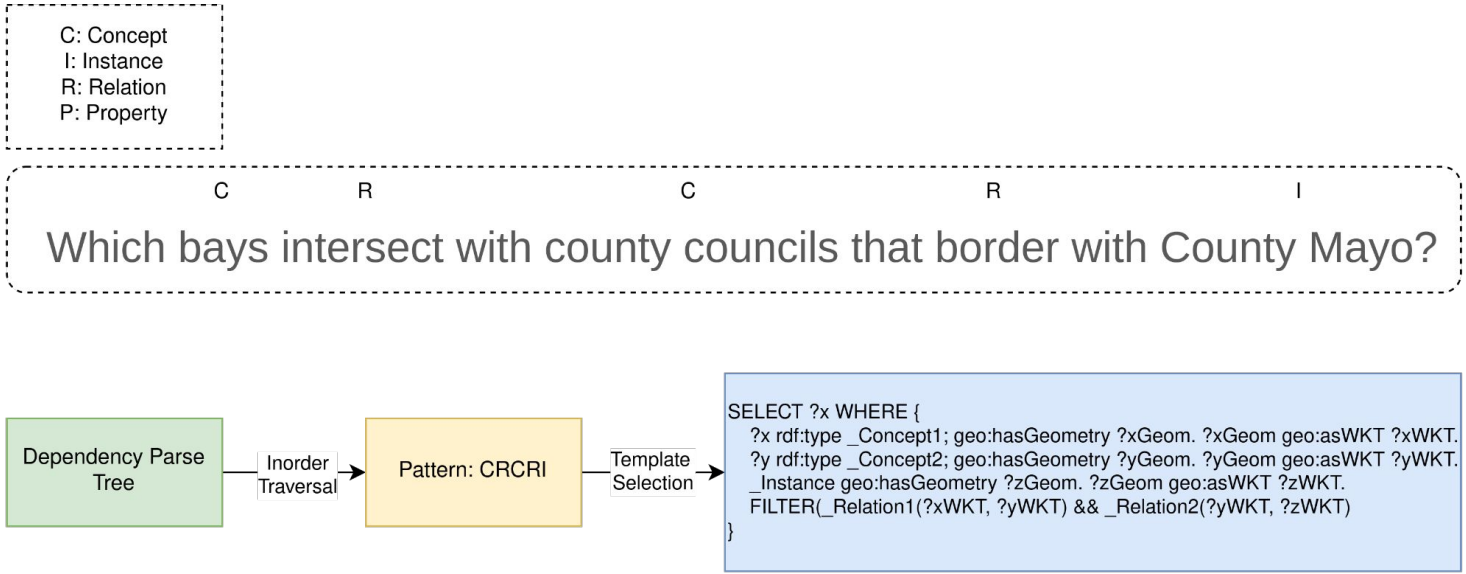


The GeoQA2 Pipeline: Running Example

- Which bays intersect with county councils that border with County Mayo?
  - Instances: yago:County_Mayo
  - Concepts: y2geoo:OSI_County_Council, y2geoo:OSM_bay
  - Relations: geof:sfIntersects, geof:sfTouches
  - Properties: -

The GeoQA2 Pipeline: Query Generator (1)

- This module generates the formal query using handcrafted query patterns, templates, and the outputs of the previous modules.



The GeoQA2 Pipeline: Query Generator (2)

- To capture more complex questions, containing **superlatives**, **comparatives**, or **aggregates**, the query generator uses the **constituency parse tree**.
- **Example**: “Which civil parishes in Ireland have more than 10 townlands?”.
  - Quantifier phrase: more than 10.
  - Modify the template to use COUNT, GROUP BY and HAVING.
- **Example**: “Which county of England has the most parks?”
  - Edge of the dependency parse tree which contains {RBS, JJS/DT} and no QP.
  - Modify the template to use COUNT, GROUP BY, ORDER BY and LIMIT.

The GeoQA2 Pipeline: Running Example

- Which bays intersect with county councils that border with County Mayo?
  - Instances: yago:County_Mayo
  - Concepts: y2geoo:OSI_County_Council, y2geoo:OSM_bay
  - Relations: geof:sfIntersects, geof:sfTouches
  - Properties: -
- Generated Query:

```
select distinct ?x where {
        ?x rdf:type y2geoo:OSM_bay; geo:hasGeometry ?cGeom1.  ?cGeom1 geo:asWKT ?cWKT1.
        ?y rdf:type y2geoo:OSI_County_Council; geo:hasGeometry ?cGeom2. ?cGeom2 geo:asWKT ?cWKT2.
        yago:County_Mayo   geo:hasGeometry ?iGeom. ?iGeom geo:asWKT ?iWKT.
        FILTER(geof:sfIntersects(?cWKT1,?cWKT2))
        FILTER(geof:sfTouches(?cWKT2,?iWKT))
}
```

Evaluation

- GeoQA2 was evaluated on the benchmark dataset GeoQuestions1089, the largest dataset containing triples of questions-queries-answers for geospatial QA.
- Answers are compared to the gold answer in the dataset and must match exactly to be counted as correct.

Category	Hamzei et al.			GeoQA2		
	Generated Queries	Correct Answers	Correct Answers*	Generated Queries	Correct Answers	Correct Answers*
Type-A	89.71%	10.85%	12.10%	84%	47.42%	56.45%
Type-B	95.68%	53.23%	55.63%	76.25%	58.99%	77.35%
Type-C	97.75%	30.33%	31.03%	79.21%	44.38%	56.02%
Type-D	100%	12%	12.00%	56%	12%	21.42%
Type-E	99.25%	7.40%	7.46%	80%	31.85%	39.81%
Type-F	79.16%	4.10%	5%	66.66%	16.66%	25%
Type-G	98.27%	11.49%	11.69%	74.13%	32.18%	43.41%
Type-H	97.18%	7.74%	7.97%	71.12%	26.05%	36.63%
Type-I	92%	0%	0.00%	84%	20%	23.80%
Total	**95.77%**	18.97%	19.81%	76.99%	**38.54%**	**50.06%**

* percentage of correct answers over the number of generated queries

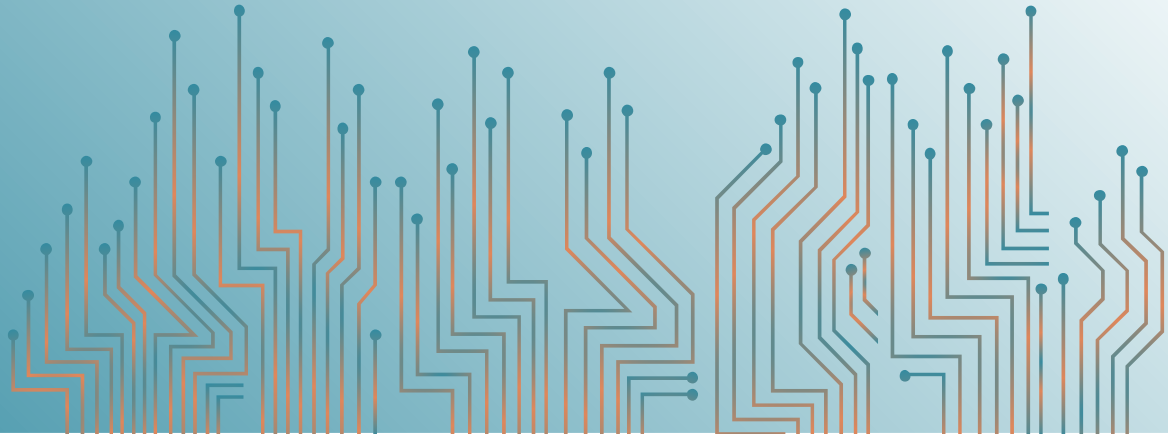
Future Work

Our group is currently working on improving GeoQA by:

- Making GeoQA able to handle spatiotemporal questions.
- Utilizing Large Language Models to improve query generation and natural language understanding.
- Utilizing state-of-the-art Entity Linking systems.

Thank you!

Find us in our web page: https://ai.di.uoa.gr/
Or follow us on twitter @AITeamUoA



Aiteam
NATIONAL & KAPODISTRIAN
UNIVERSITY OF ATHENS