Open Domain Question Answering over Knowledge Graphs using Keyword Search, Answer Type Prediction, SPARQL and Pre-trained Neural Models

Christos Nikas, Pavlos Fafalios and Yannis Tzitzikas

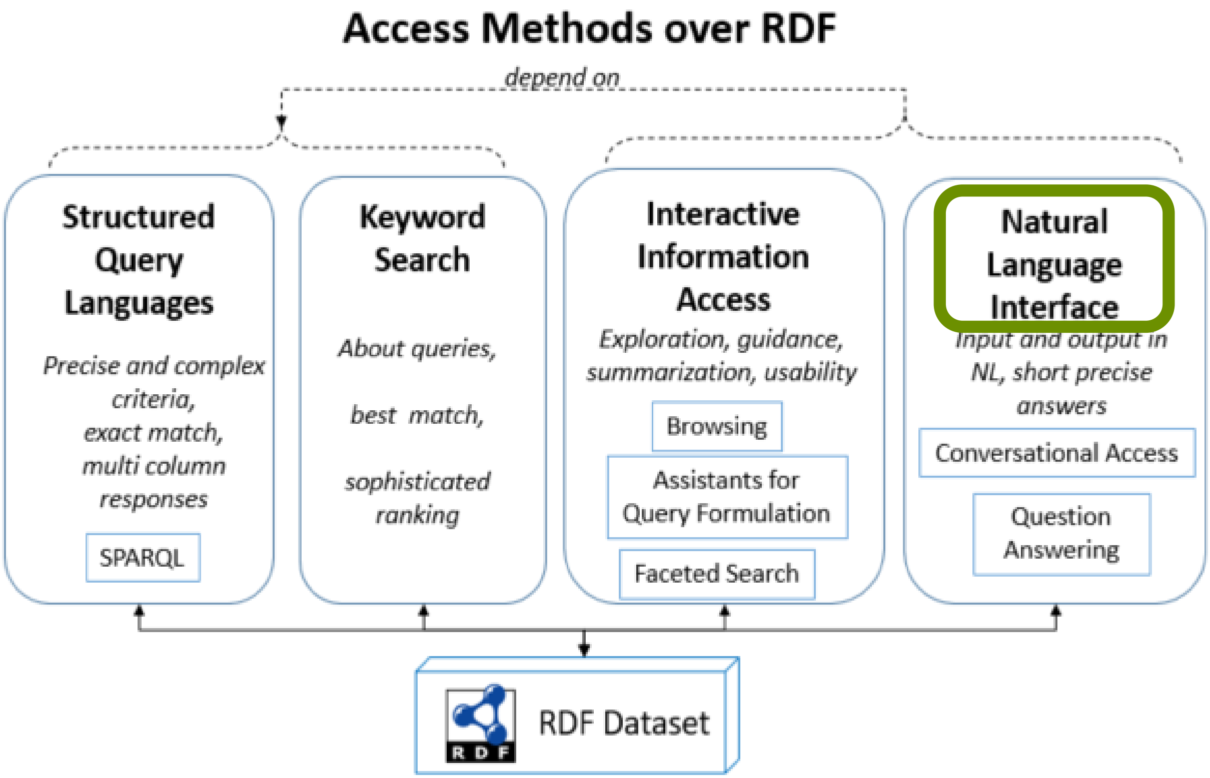
Information Systems Laboratory, FORTH-ICS, and
Computer Science Department, University of Crete, Greece

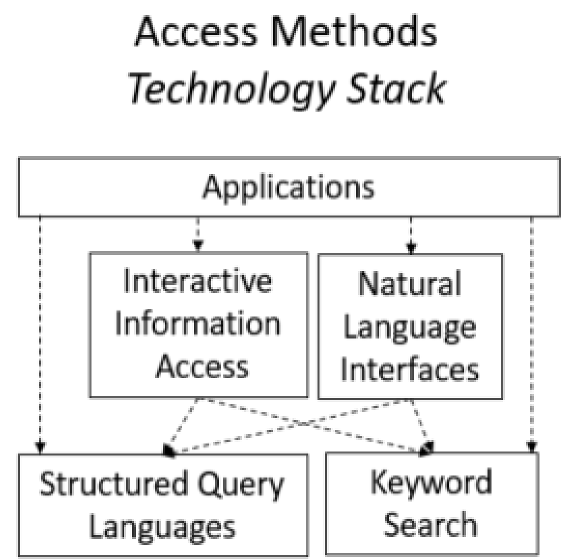
ISWC 2021

Outline

- Context & Motivation
- Overview of the Approach
- Steps of the QA pipeline
- Evaluation
  - Research Questions
  - Evaluation Results
- Concluding Remarks

Context: Natural Language Interfaces to RDF Datasets





Motivation and Approach Overview

Problem:

- QA in open domain information needs is hard to be adequate, satisfying and pleasing for end users

Approach

- We investigate an approach where QA complements a general purpose interactive keyword search over RDF.

Rationale

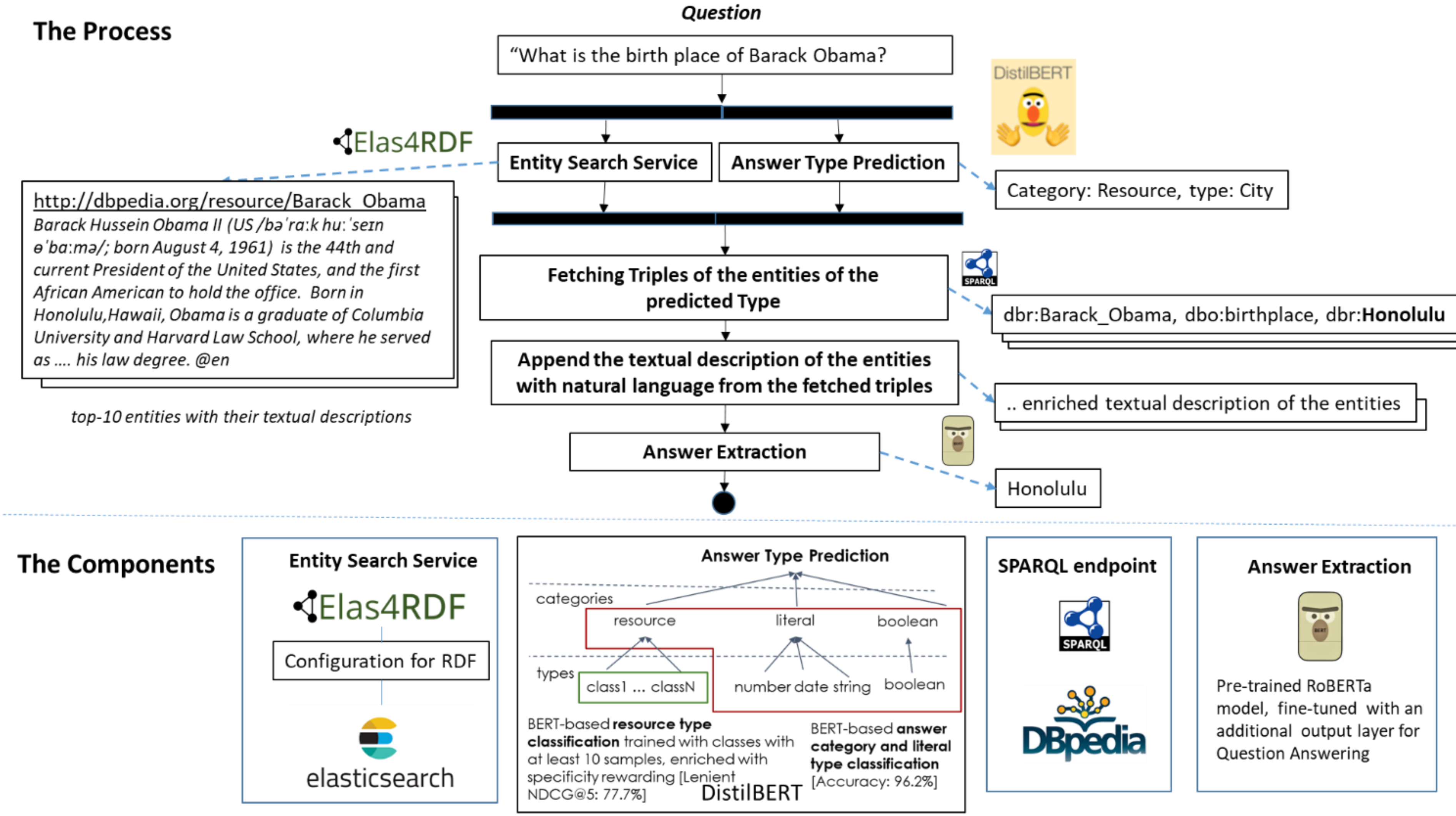
- Having **a keyword search** service at its core, promises various benefits:
  - Allows exploiting the wealth of techniques related to text pre-processing, retrieval and language models, thus tackling some of the weaknesses of current components
    - related to the upper/lowercase of named entities, the implicit entity names (that NER tools usually fail to identify due to the various morphological variations), the abbreviations in named entities, and others.
  - Not all question intentions can be identified and mapped to the correct SPARQL statement (e.g. questions that can be answered by the textual descriptions in the rdfs:comment)

Investigated QA Pipeline

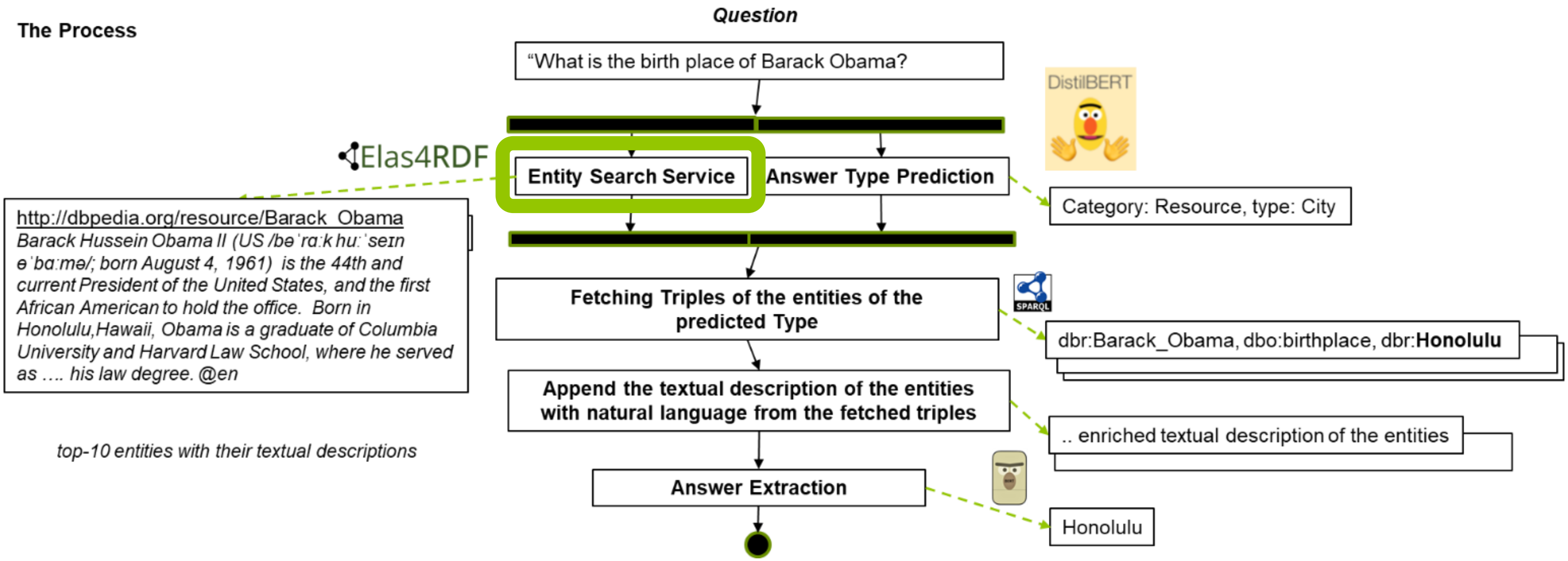
- A pipeline for open domain QA comprising Keyword Search, Answer Type Prediction, Entity Expansion and Answer Extraction

FORTH INSTITUTE OF COMPUTER SCIENCE ISL

Overview of the Approach



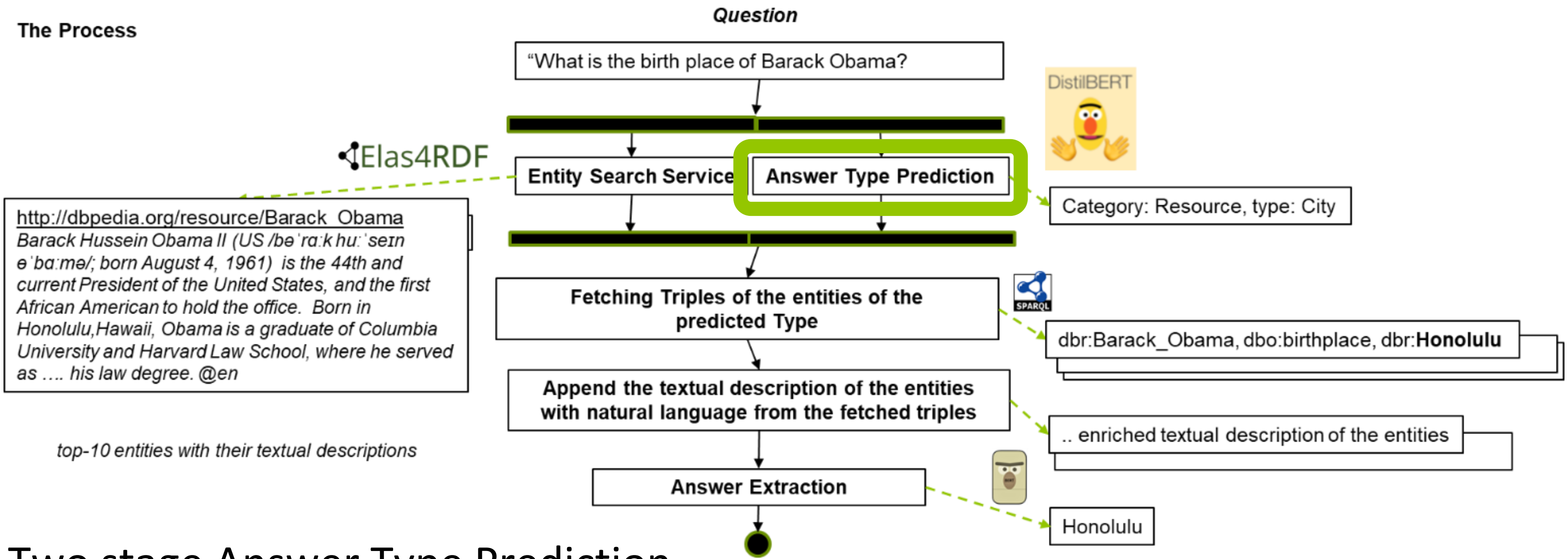
Task: Entity Search



- We leverage the Elas4RDF approach for keyword search over RDF (Kadilierakis et al, ESWC'2020)
  - A triple-centered approach
  - G. Kadilierakis, P. Fafalios, P. Papadakos, Y. Tzitzikas, Keyword search over RDF using Document-Centric Information Retrieval Systems, Proceedings of the 17th Extended Semantic Web Conference (ESWC'2020), June 2020, Heraklion, Crete

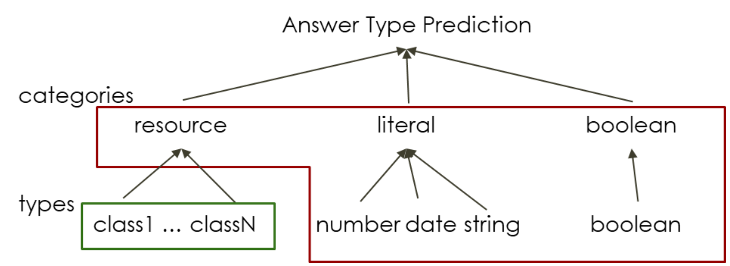
FORTH INSTITUTE OF COMPUTER SCIENCE

Task: Answer Type Prediction



Two stage Answer Type Prediction

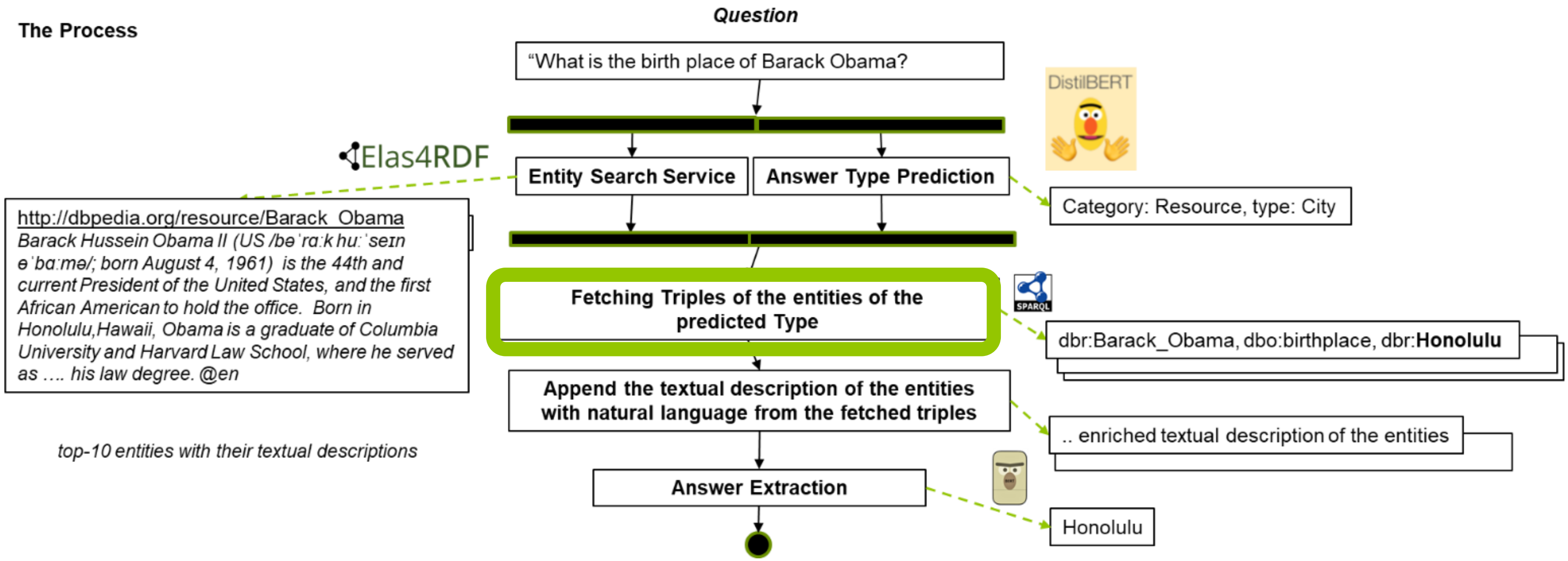
1. Category Prediction (Boolean, Literal, Resource)
2. Type Prediction
   - Literal questions (Date, Number, String)
   - Resource questions (a class in the DBpedia ontology)



SeMantic AnsweR Type prediction task (https://smart-task.github.io/)

- ❑ Dataset including questions and answer types from DBpedia and Wikidata
- ❑ Challenge in ISWC 2020
- ❑ Our approach gained 2nd place

Task: Fetch related triples



- We use SPARQL queries to fetch facts about the retrieved entities that have the expected type.

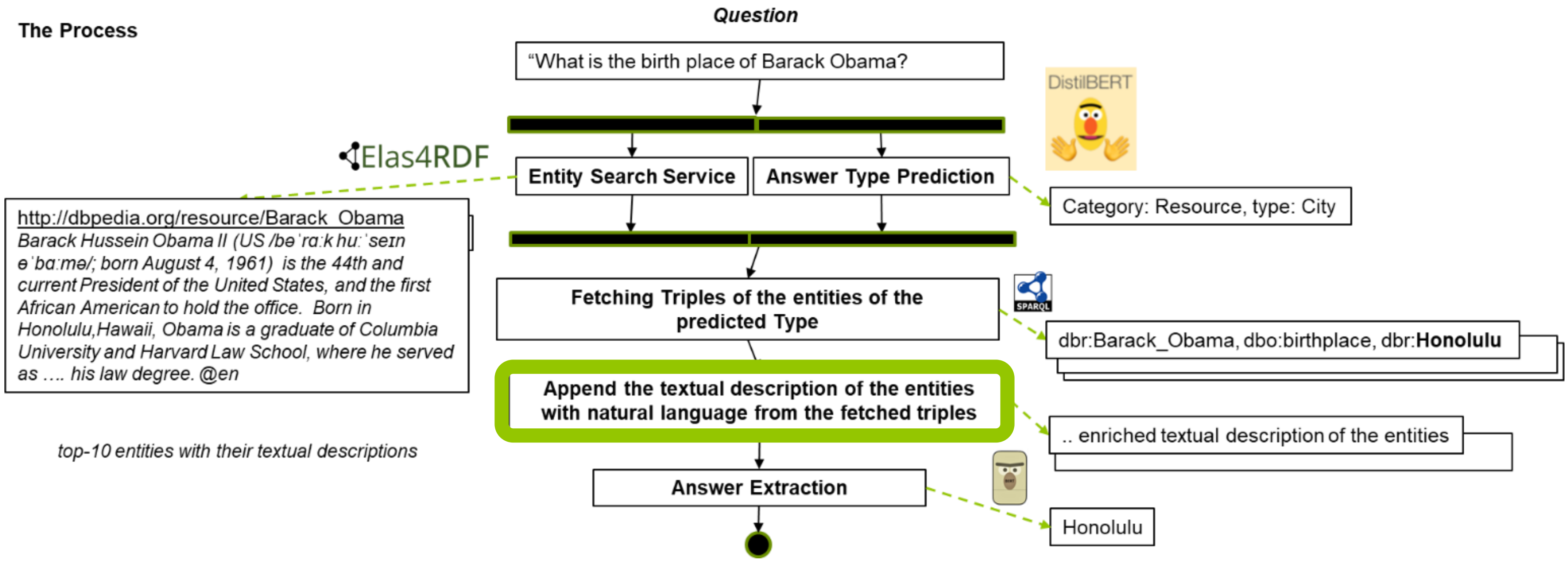
Resource

```
select distinct str(?pl) as ?pLabel ?a where {
    <entity uri> ?p ?a .
    ?p rdfs:label ?pl .
    <answer type> owl:equivalentClass ?eq .
    ?a rdf:type ?eq .
    FILTER(lang(?pl) = 'en' || lang(?pl) = '')
}
```

Literal

```
select str(?answer) as ?a str(?pl) as
?pLabel where {
     <entity uri> ?p ?answer .
     ?p rdfs:range <xsd:type> .
     ?p rdfs:label ?pl .
          FILTER(isLiteral(?answer))
     FILTER(lang(?pl) = 'en' ||
     lang(?pl) = '')
}
```

Task: Entity description enrichment



- We generate sentences from the facts retrieved by SPARQL and we append them to the entity descriptions
- Example (over running) on next slide.

Task: Entity description enrichment

- Textual description before:

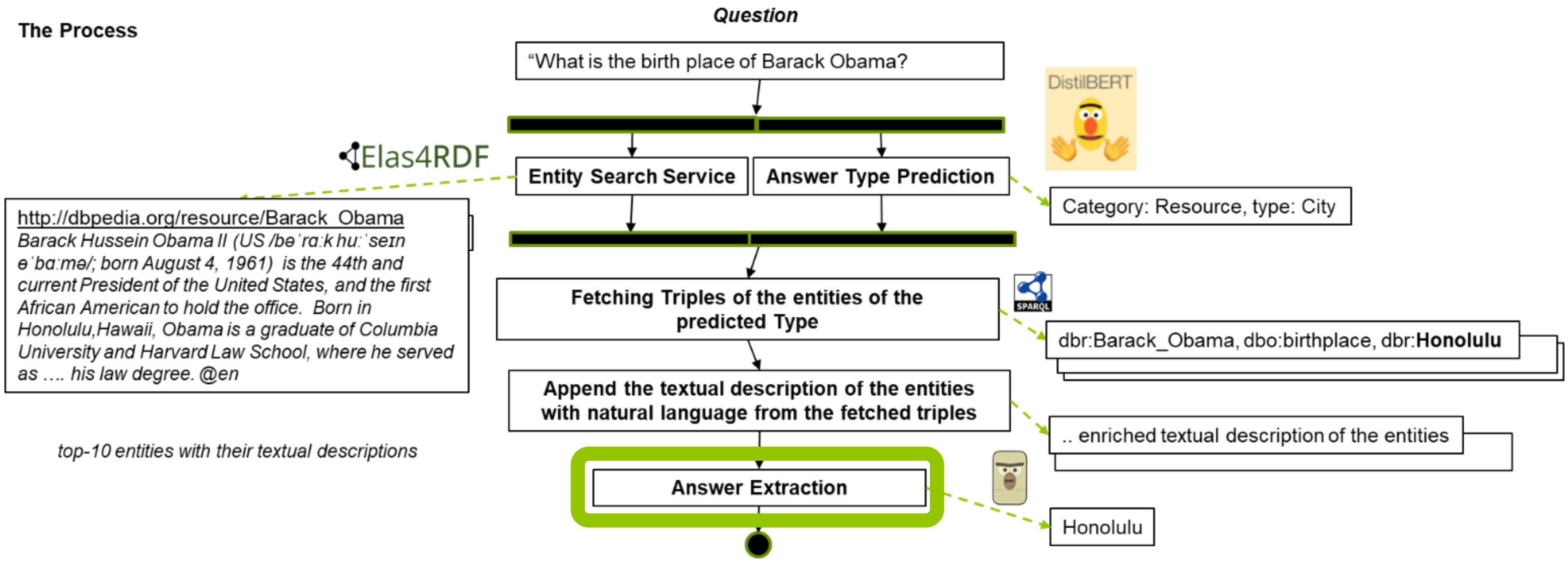
Barack Hussein Obama II (US /bəˈrɑːk huːˈseɪn ɵˈbɑːmə/; born August 4, 1961) is the 44th and current President ... He was a community organizer in Chicago before earning his law degree.

Retrieved RDF triple	dbr:Barack_Obama, dbo:birthplace, dbr:Honolulu
generated sentence	Barack Obama birth place Honolulu

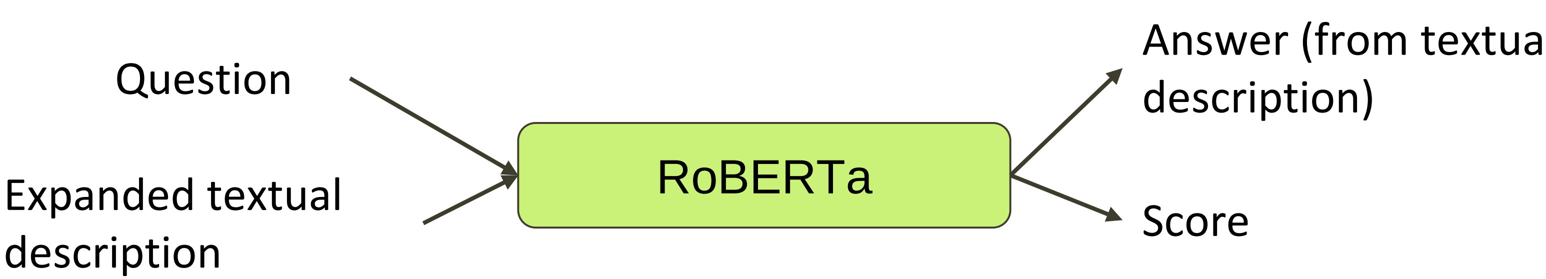
- Textual description after:

Barack Hussein Obama II (US /bəˈrɑːk huːˈseɪn ɵˈbɑːmə/; born August 4, 1961) is the 44th and current President ... He was a community organizer in Chicago before earning his law degree. Barack Obama birth place Honolulu.

Task: Answer Extraction



- We use RoBERTa (a variation of BERT fine-tuned using more data and computing resources than BERT; more powerful)



A Key point

Compared to other systems that obtain high scores over open domain QA, our system:

- Does not follow a supervised end-to-end approach, trained on the same knowledge base
- Makes use of **different information sources** than those intended by the benchmarks.

We evaluate how good our approach for open domain QA is while retrieving information from a **different source** and without having been previously trained over this specific dataset.

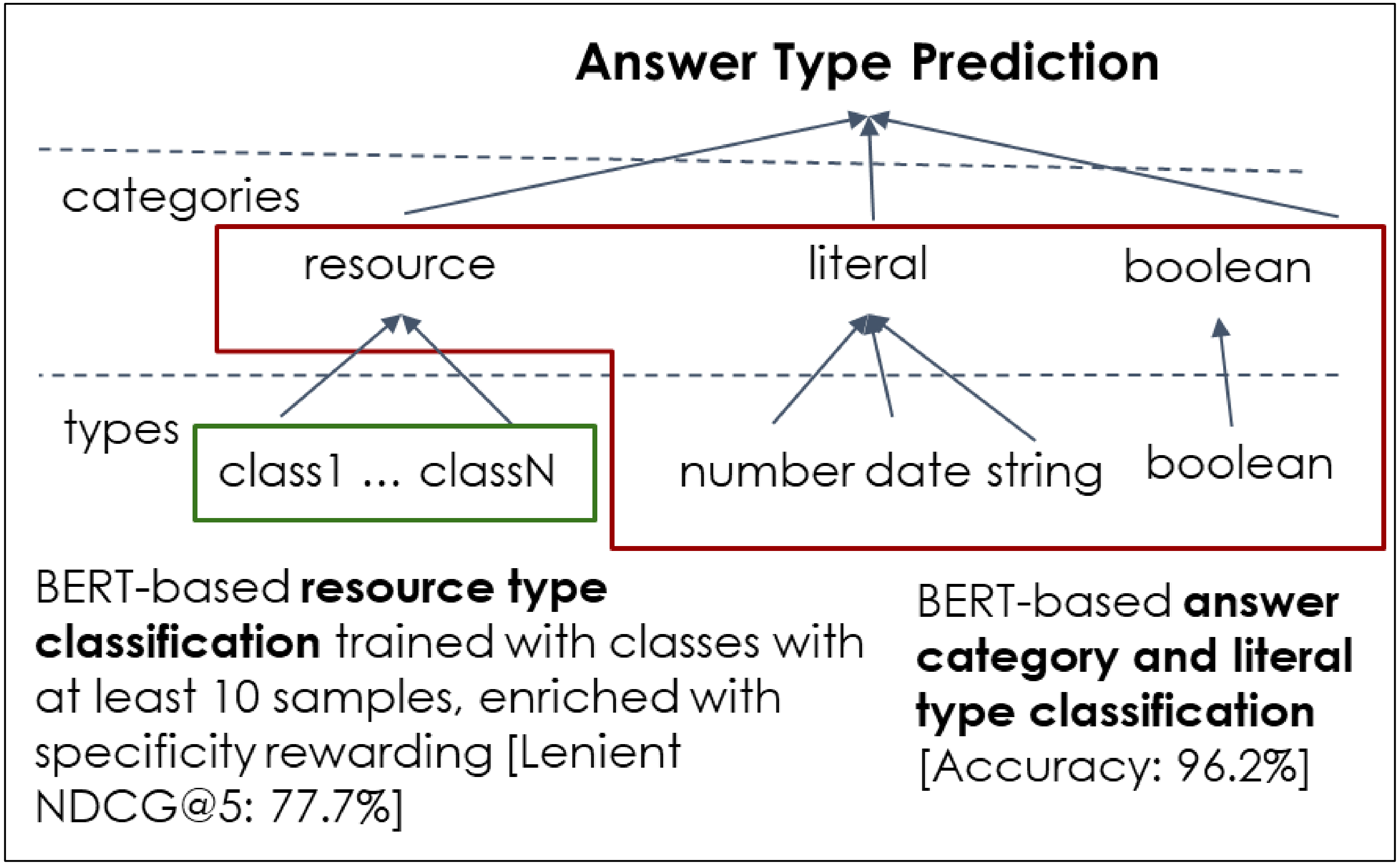
FORTH INSTITUTE OF COMPUTER SCIENCE isl

Related Research Questions

- (a) How effective is the Answer Type Prediction?
- (b) How good can the QA pipeline over DBpedia be, in comparison to approaches and benchmarks over a different knowledge graph (in our case Freebase)?
- (c) How does Answer Type Prediction affect the quality of QA?
- (d) How can answers from this QA pipeline contribute to the entity retrieval task over DBpedia-Entity dataset [7], and entity ranking in general?

FORTH INSTITUTE OF COMPUTER SCIENCE

Evaluation of Answer Type Prediction



Christos Nikas, Pavlos Fafalios and Yannis Tzitzikas,
[Two-stage Semantic Answer Type Prediction for Question Answering using BERT and Class-Specificity Rewarding](),
SeMantic AnsweR Type prediction (SMART) Challenge, 2020 International Semantic Web Conference (ISWC'2020) Challenge, Nov 2020.

Evaluation of QA over WebQuestions

We answer all questions in the test collection and obtain the following results:

Threshold	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Precision	7	16.12	18.6	21.29	23.71	25.26	28.10	31.19	37.54	**43.36**
Recall	31.27	27.71	29.02	27.89	29.08	30.88	31.28	33.47	34.51	**40.48**
F1	9.71	16.96	19.07	19.7	21.66	23.22	25.04	28.36	31.44	**39.20**
Accuracy	**53.76**	47.6	47.6	46.9	47.69	48.03	47.87	48.77	52.38	52.17

Results for varying answer score threshold

FORTH INSTITUTE OF COMPUTER SCIENCE ISL

Evaluation of QA over WebQuestions: **without Answer Type Prediction**

- We perform the same experiment without Answer Type Prediction. Entity descriptions are not expanded
- For the best value for answer score threshold (0.9)

	with ATP	without ATP
Precision	43.36	37.36
Recall	40.48	32.97
F1	39.20	32.18
Accuracy	52.17	48.12

Answer Type Prediction + Entity Expansion improves results by 4-8% (relative improvement 11%-17%)

FORTH INSTITUTE OF COMPUTER SCIENCE ISL

Evaluation over DBpedia entity QA

We answer all questions in the QA subset of DBpedia Entity

Threshold	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
P@1	33.57	49.33	55.43	55.64	55.94	56.19	58.86	57.24	57.27	**69.44**
P@3	27.84	42	48.27	47.95	46.6	50.81	52.91	51.21	52.35	**69.44**
P@5	24.54	41.01	47.15	46.84	45.77	49.77	51.91	51.21	52.35	**69.44**

Precision @1, @3, @5 for varying answer score threshold over DBpedia Entity

DBpedia entity QA + Ranking

- We evaluate our system for the task of Entity Search
- We compute NDCG@10,100 over the QA subset of the DBpedia Entity collection using:
  - Entities returned from the Elas4RDF Search Service
  - Entities returned from the Elas4RDF Search Service + Top answers from the QA component

"How can the QA component improve Entity Search tasks?"

DBpedia entity QA + Ranking

Approach 1:
Keep the score from each entity and answer as computed by the entity search system and question answering component

	NDCG@100		NDCG@10	
Answers added	Score	Difference	Score	Difference
0 (baseline)	0.325	0	0.325	0
1	0.352	0.027	0.352	0.027
3	0.372	0.047	0.353	0.028
5	**0.384**	**0.059**	**0.354**	**0.029**
10	0.382	0.057	0.353	0.028

Approach 2:
Sum scores for entities in both rankings

	NDCG@100		NDCG@10	
Answers added	Score	Difference	Score	Difference
0 (baseline)	0.325	0	0.325	0
1	0.355	0.03	0.355	0.03
3	0.375	0.05	**0.358**	**0.033**
5	**0.387**	**0.062**	0.357	0.032
10	0.386	0.061	0.356	0.031

Efficiency

Answer Type Prediction	Entity Expansion	Answer Extraction
0.1 sec (1.2%)	3.9 sec (47%)	4.3 sec (51.8%)

- memory footprint: ~1.4 GB
- required space: ~511 MB (for all models)

Application: https://demos.isl.ics.forth.gr/elas4rdf/

Attempts to interpret the query as a question, and find triples that contain a natural language answer

Elas4RDF who is the father of Queen Elizabeth II?

Triples Entities Graph Schema QA

Answer Type

Category

resource

Type

Person

Answer category and semantic type

Answers

Entity URI

Entity description

King George VI,
From entity:
http://dbpedia.org/resource/Coronation_of_Queen_Elizabeth_II
The coronation of Queen Elizabeth II as monarch of the United Kingdom, Canada, Australia, New Zealand, Union of Sout
on 2 June 1953. Elizabeth ascended the thrones of these countries at age 25, upon the death of her father, **King George**
proclaimed queen by her various privy and executive councils shortly afterwards.
Score: 0.856

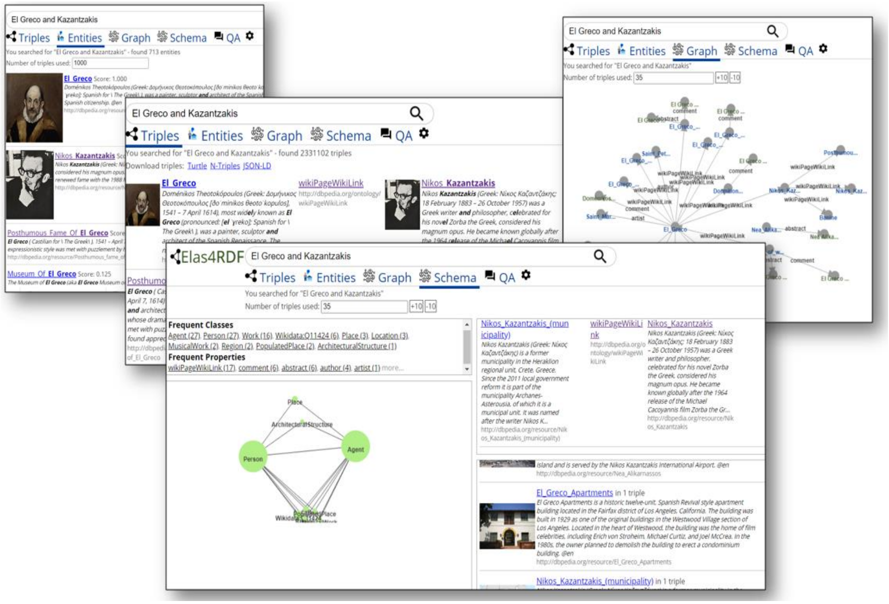
Prince Charles,
From entity:
http://dbpedia.org/resource/Monarchy_of_Belize
The monarchy of Belize (the Belizean monarchy) is a system of government in which a hereditary monarch is the sovere
Elizabeth II, officially called Queen of Belize, who has reigned since 21 September 1981. The heir apparent is Elizabeth's
Queen is the only member of the royal family with any constitutional role. Monarchy of Belize first monarch Elizabeth II.
Charles, Prince of Wales. Monarchy of Belize incumbent Elizabeth II
Score: 0.592

FORTH INSTITUTE OF COMPUTER SCIENCE isl

Concluding Remarks

- We have proposed an approach for open domain QA that obtains satisfactory results, i.e. 54% accuracy, 39% F1 over popular QA benchmarks, something that is very interesting because **it does not follow a supervised end-to-end approach trained on the same knowledge base**, but makes use of **different information sources than those intended by the benchmarks!**
- Answer Type Prediction and Entity Enrichment stages improve Precision by 6%, Recall by 7% and F1 score by 7% (over WebQuestions).
- The proposed approach can be used in combination with an entity search system to improve entity search tasks by 6% NDCG@100 (over DBpedia Entity dataset).
- Overall, the proposed pipeline can be applied over large knowledge graphs, since the process starts from an efficient and effective keyword search system, while the next steps exploit pre-trained neural network models.

Thanks for your attention



Online demo:

https://demos.isl.ics.forth.gr/elas4rdf/

Preprint of the full paper:

http://users.ics.forth.gr/~tzitzik/publications/Tzitzikas_2021_ISWC-QA.pdf

FORTH INSTITUTE OF COMPUTER SCIENCE