QALD-7
Question Answering over Linked Data Challenge

Presenter: Giulio Napolitano

QALD-7 @ ESWC 2017
Portoroz, Slovenia

HOBBIT
Holistic Benchmarking of Big Linked Data

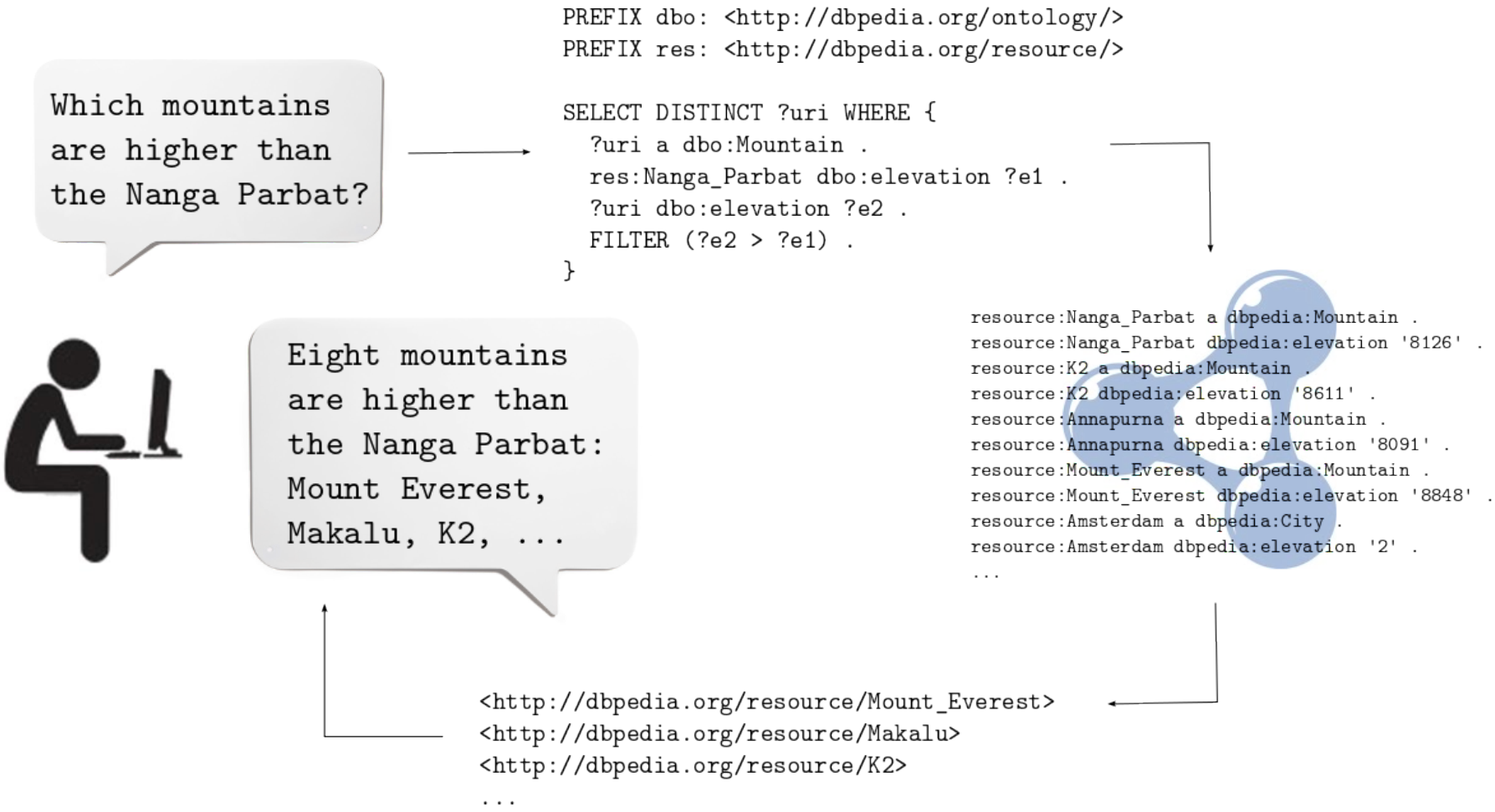
Horizon 2020, GA No 688227

May 30th, 2017

Question answering systems mediate between

a user expressing an information need in *natural language*

and *RDF-modelled* data



QALD is a series of evaluation campaigns that provide a benchmark for comparing different approaches and systems

- get a picture of their strengths and shortcomings
- gain insight into how we can develop approaches that deal with Semantic Web data as a knowledge source

QALD-1 @ ESWC 2011 (3)

QALD-2 @ ESWC 2012 (4)

QALD-3 @ CLEF 2013 (6)

QALD-4 @ CLEF 2014 QA track (9)

QALD-5 @ CLEF 2015 QA track (7)

QALD-6 @ ESWC 2016 (13)

QALD-7 @ ESWC 2017 (3)

Overall task Given a natural language question, retrieve the correct answer(s) from a given RDF repository.

Types of challenges (specific tasks):

- Multilingual
- Hybrid
- Large scale
- Wikidata

Task 1 - Multilingual questions

Dataset: DBpedia 2016-04 (with multilingual labels)

Questions: 215 training, 50 test

- provided in 8 languages: English, German, Spanish, Italian, French, Dutch, Romanian, Farsi
- can be answered with respect to the provided RDF data
- annotated with corresponding SPARQL queries and answers

Challenge: Lexical and structural gap between natural language expressions and data, e.g.

- high $\rightarrow$ `elevation`
- have inhabitants $\rightarrow$ `populationTotal`
- graduate from $\rightarrow$ `almaMater`

Example

Which book has the most pages?

Welches Buch hat die meisten Seiten?

Quale libro ha il maggior numero di pagine?

Quel livre a le plus de pages?

¿Que libro tiene el mayor numero de paginas?

...

Dataset: DBpedia 2016-04 (with free text abstracts)

Questions: 105 training, 50 test

provided in English

can be answered only by integrating structured data (RDF) and unstructured data (free text abstracts)

annotated with pseudo-queries and answers

Challenge: find information in several sources, process both structured and unstructured information, and combine them into one answer.

Example

Who is the front man of the band that wrote Coffee & TV?

HOBBIT
Holistic Benchmarking of Big Linked Data

```
PREFIX res: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
SELECT DISTINCT ?uri
WHERE {
        res:Coffee_&_TV dbo:musicalArtist ?x .
        ?x dbo:bandMember ?uri .
        ?uri text:"is" text:"frontman" .
}
```

`http://dbpedia.org/resource/Damon_Albarn`

Dataset: DBpedia 2016-04

Questions: 100 training, 2M test

provided in English

automatically generated

questions sent every minute, *n+1* questions asked at minute *n*

Challenge: deal with high volume requests in a short time

Dataset: Wikidata 2017-01

Questions: 100 training, 50 test

provided in English

questions based on DBpedia but performed on Wikidata

Challenge: formulate generic approaches, adapting to new data sources

Dennis Diefenbach, Kamal Singh, Pierre Maret
WDAqua-core0: A Question Answering Component for the Research Community
Task 1 and Task 4

Nikolay Radoev, Mathieu Tremblay, Michel Gagnon, Amal Zouaq
Answering Natural Language Questions on RDF Knowledge base in French
Task 1

Daniil Sorokin, Iryna Gurevych
End-to-end Representation Learning for Question Answering with Weak Supervision
Task 4

Organization committee

Ricardo Usbeck
Universität Leipzig, Germany

Axel-Cyrille Ngonga Ngomo
Universität Leipzig, Germany

Bastian Haarmann
Fraunhofer Institute IAIS, Germany

Anastasia Krithara
National Center for Scientic Research “Demokritos", Greece

Harsh Takkar
Universität Bonn, Germany

Henning Petzka
Fraunhofer Institute IAIS, Germany

Jens Jehmann
Fraunhofer Institute IAIS, Germany

Program committee

HOBBIT Holistic Benchmarking of Big Linked Data

Corina Forascu - Alexandru Ioan Cuza University, Iasi, Romania

Sebastian Walter - CITEC, Universität Bielefeld, Germany

Bernd Müller - ZBMed, Germany

Christoph Lange - Fraunhofer Gesellschaft, Germany

Dennis Diefenbach - Université de Saint-Étienne, France

Edgard Marx - Universität Leipzig, Germany

Hady Elsahar - Université de Saint-Étienne, France

Ioanna Lytra - Universität Bonn, Germany

John McCrae - INSIGHT - The Centre for Data Analytics, Ireland

Konrad Höffner - Universität Leipzig, Germany

Kuldeep Singh - Universität Bonn, Germany

Saeedeh Shekarpour - Kno.e.sis Center, Ohio Center of Excellence in Knowledge-enabled Computing, USA

Sherzod Hakimov - CITEC, Universität Bielefeld, Germany

Thank You!

Thanks to ***Christina Unger*** for sharing her slides!!!!