

WDAqua: Answering Questions using Web Data

Ioanna Lytra¹, Maria-Esther Vidal^{1,2}, Christoph Lange^{1,2}, Elena Demidova³,
and Sören Auer^{1,2}

¹ University of Bonn, Germany `lytra,vidal,langec,auer@cs.uni-bonn.de`

² Fraunhofer IAIS, Germany . . .

³ University of Southampton, UK `e.demidova@soton.ac.uk`

1 WDAqua in a Nutshell

WDAqua, a Marie Skłodowska Curie Innovative Training Network (ITN) running from January 2015 to December 2018, involves six academic partners⁴, and employs 15 PhD students (early stage researchers, ESRs) in total. The main motivation of this project is that sharing, connecting, analyzing, and understanding data on the Web can provide better services to citizens, communities, and the industry. A vehicle to achieve this is data-driven question answering (QA), having the key objective of delivering precise and comprehensive answers to natural language questions primarily by making better use of data. Powerful QA tools promise to improve access to the large amount of information available on the Web, or even private data collections, and can be immediately useful to a wide audience of end users in their private and professional life. Data-driven QA comprises four simplified steps: 1) understanding a human question, and turning it into natural language text, 2) analyzing the question in natural language, 3) finding data to answer the question and to justify the answer, and finally, 4) presenting the answer using, e.g., verbalization, natural language synthesis, or visualization.

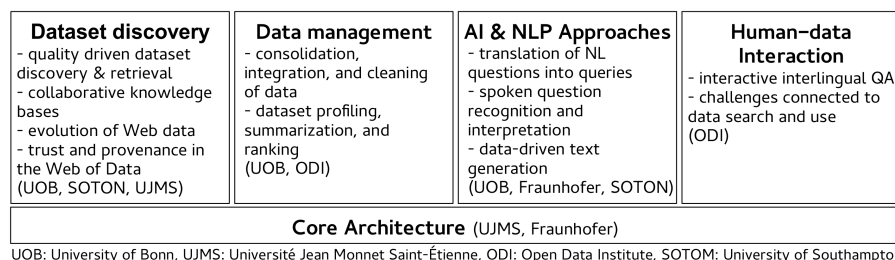
2 Project Progress and Dissemination

The aim of the WDAqua project is to advance the state of the art in the challenging research field of data-driven QA by interleaving training, research, and innovation. At the final stage of the project, the respective research results will be combined into an open data-driven QA platform and ecosystem, where other QA related components from the research community can be plugged in. The recruitment of the ESRs has now been completed and various research and training events have been organized by the WDAqua members. In the 1st WDAqua Learning Week the ESRs gained hands-on experience on different components of QA systems and to establish collaborations with fellow WDAqua ESRs. In the

⁴ University of Bonn (coordinator), Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS, Germany; National and Kapodistrian University of Athens, Greece; Université Jean Monnet Saint-Étienne, France; University of Southampton, Open Data Institute (ODI), UK.

1st R&D week, WDAqua ESRs strengthened their collaborations and initiated several ongoing research projects on topics related to Linked Data, multilingual data analytics and Q&A system design. Upcoming training activities include the ESWC 2016 summer school on Data Science in September 2016⁵.

At this early stage of the WDAqua research project, the researchers focused, first of all, on identifying and investigating open research questions in data-driven QA, analyzed the state of the art regarding these open challenges, and published or submitted for publication some initial results at high-quality research venues (ICSC 2016, ESWC 2016, etc.). The research questions the PhD students are currently working on can be classified into four key research areas strongly related to data-driven QA (cf. Fig. 1): dataset discovery (i.e., quality-driven dataset discovery and retrieval, research on collaborative knowledge bases to support question answering, handling evolution of Web data through extraction of facts in free text, trust and provenance in the Web of data), data management (making datasets fit for QA, dataset profiling, summarization, and ranking), AI and NLP approaches (translating natural language questions in queries, spoken question recognition and interpretation, data-driven text generation), and human-data interaction (interactive interlingual QA systems, challenges connected to data search and use). Apart from that, the project is concerned with designing the architecture of an open QA platform, to enable the integration of the different QA components that will be developed by the partners of WDAqua and beyond.



UOB: University of Bonn, UJMS: Université Jean Monnet Saint-Étienne, ODI: Open Data Institute, SOTOM: University of Southampton

Fig. 1. Research challenges related to data-driven QA

3 Networking

The WDAqua ITN aims at reaching out to the scientific community, industry, and society. The first steps in pursuing challenging research questions in this promising research area have been conducted and efforts have been invested in attracting input by the community, by presenting WDAqua at European and international research venues, as well as in transferring expertise to and exchanging experience with the research community by organizing training events (i.e., summer schools, learning weeks) and establishing collaborations with the industry (e.g., through internships of PhD students).

⁵ <http://summerschool2016.eswc-conferences.org>