SIGIR 2017 Workshop on
Open Knowledge Base and Question Answering (OKBQA2017)

Key-Sun Choi
KAIST
kschoi@kaist.ac.kr

Teruko Mitamura
CMU
teruko@cs.cmu.edu

Piek Vossen
Vrije Universiteit Amsterdam
p.vossen@let.vu.nl

Jin-Dong Kim
Database Center for Life Science
jdkim@dbcls.rois.ac.jp

Axel-Cyrille Ngonga Ngomo
University of Paderborn
axel.ngonga@upb.de

ABSTRACT

Over the past years, several challenges and calls for research projects have pointed out the dire need for pushing natural language interfaces. In this context, the importance of Semantic Web data as a premier knowledge source is rapidly increasing. But we are still far from having accurate natural language interfaces that allow handling complex information needs in a user-centric and highly performant manner. The development of such interfaces requires collaboration of a range of different fields, including natural language processing, information extraction, knowledge base construction and population, reasoning, and question answering. With the goal to join forces in the collaborative development of natural language QA systems, the second OKBQA workshop is organized within the 40th SIGIR conference.

1 INTRODUCTION

The huge and rapidly increasing amount of structured and unstructured data available on the Web makes it both possible and necessary to support users in finding relevant information. The trend moves more and more towards smart knowledge services that are able to find information, aggregate them, draw inferences, and present succinct answers without requiring the user to wade through a large number of documents. The novel avenues made possible by knowledge services are numerous and diverse, including ubiquitous information access (from smartphones, tablets, smart watches, etc.), barrier-free access to data (especially for the blind and disabled) and knowledge discovery.

Over the past years, several challenges and calls for research projects have pointed out the dire need for pushing natural language interfaces. In this context, the importance of Semantic Web

data as a premier knowledge source is rapidly increasing. But we are still far from having accurate natural language interfaces that allow handling complex information needs in a user-centric and highly performant manner. The development of such interfaces requires the collaboration of a range of different fields, including natural language processing, information extraction, knowledge base construction and population, reasoning, and question answering.

The main goal of this workshop is to join forces in the collaborative development of open frameworks for knowledge extraction and question answering, to share standards, and to foster the creation of an ecosystem of tools and benchmarks. The workshop will therefore not only comprise short and long paper presentations but also a hands-on session on already existing frameworks, standards, and benchmarking campaigns, as well as a social meet-up.

2 THEME AND PURPOSE OF THE WORKSHOP

Development of natural language QA systems requires expertise in various technologies such as natural language processing, information extraction, knowledge base construction and population, reasoning, and so on. The OKBQA community has been working on establishing an open collaboration among interested parties in developing natural language interfaces to knowledge bases. The OKBQA hackathon has been held four times so far, and the OKBQA platform (http://doc.okbqa.org) is developed as a main product. The OKBQA workshop is a new series of forum for exchanging frontier works relevant to natural language QA and knowledge bases. The first OKBQA workshop was held in Osaka, Japan, within COLING 2016 (International Conference on Computational Linguistics) (http://coling2016.okbqa.org/), and we met many participants from NLP community. The second workshop is organized within SIGIR 2017, for an intention to widen the community by soliciting interested scholars and engineers from the general IR community.

2.1 Topics of interest

The topic of interest to be discussed during the workshop ranges over natural language QA, KB construction and NLP annotation framework for QA, as listed below:

- Natural Language Interface for Web of Data
  - Browsing Linked Data
  - Question Answering over Linked Data
  - Benchmarking Natural Language Interfaces

- – Term Matching and Entity Disambiguation
- – SPARQL Query Pattern Generation
- – Schema-agnostic Query Generation
- – Discovery of Linked Data Sources
- – Endpoint Profiling
- – Dealing with Data and Schema Heterogeneity
- – Providing Justifications of Answers and Conveying Trust
- – Knowledge Base Design for Question Answering
- – NLP resources for Question Answering
- – Reasoning for Question Answering
- – User Feedback and Interaction
- – Dialogue Systems
- – Personal Assistants
- Knowledge Base Construction for Question Answering
  - – Entity-centered Knowledge Bases
  - – Event-centered Knowledge Bases
  - – Human Intervention of Knowledge with QA
  - – NLP Annotations for Knowledge Extraction and Machine Reading
  - – Gold Standard Data Sets and Quality Assessment
  - – Indexing and Mappings from Existing Sources (Structured, Semi-structured or Unstructured)
- NLP Annotation Framework for Knowledge Base and Question Answering
  - – NLP annotation framework - to represent the layers and their structure of language understanding in graph formats such as XML and RDF, including metadata and underlying ontologies
  - – NLP annotation knowledge base - to access existing annotated NLP resources efficiently and effectively for further processing for knowledge base engineering and question answering

3 WORKSHOP DESIGN

OKBQA'17 is a full day workshop. As well as oral presentations and poster presentations of selected papers, the workshop program will include a joint panel discussion with KG4IR workshop, held in paralell at the SIGIR 2017. KG4IR group will bring the theme of Knowledge Graphs and Semantics for Text Retrieval and Analysis, whereas OKBQA group will bring the theme of Open Knowledge Base and Question Answering. Dispite some similarities, these two workshops have a different focus. It would be interesting to get together to discuss different perspectives in the research areas.

With the paper presentations and the joint panel discussions, we aim at providing a forum for exchange of not only ideas and theories but also actual resources for engineering.

4 ORGANISATION

The organizing committee includes active members in the area of NLP, QA, Semantic Web and KB construction, who are the authors of this abstract. The program committee members support the workshop by providing review for screening of submitted papers, and includes the following members:

- André Freitas (University of Passau, Germany)
- Christina Unger (Universität Bielefeld, Germany)
- Eun-Kyung Kim (KAIST, Korea)
- Pum-mo Ryu (Busan University of Foreign Studies, Korea)
- Ricardo Usbeck (University of Leipzig, Germany)
- Jun Araki (Carnegie Mellon University, USA)
- Peter Clark (Allen Institute for AI, USA)
- Eduard Hovy (Carnegie Mellon University, USA)
- Madoka Ishioroshi (National Institute of Informatics, Japan)
- Hiroshi Kanayama (IBM Research - Tokyo, IBM Japan)
- Bernardo Magnini (Fondazione Bruno Kessler, Italy)
- Tatsunori Mori (Yokohama National University, Japan)
- Yuta Nakashima (Osaka University, Japan)
- Eric Nyberg (Carnegie Mellon University, USA)
- Anselmo Peñas (UNED, Spain)
- John M. Prager (IBM T.J. Watson Research Center, USA)
- Kotaro Sakamoto (Yokohama National University, Japan)
- Hideyuki Shibuki (Yokohama National University, Japan)
- Hideki Shima (Duolingo, Inc., USA)
- Koichi Takeda (IBM Research - Tokyo, IBM Japan)
- Chuan-Jie Lin (National Taiwan Ocean University, Taiwan)
- Di Wang (Carnegie Mellon University, USA)
- Luke Zettlemoyer (University of Washington, USA)

5 ENDORSEMENT

SIGIR (Special Interest Group on Information Retrieval) provided their endorsement to the workshop due to the relevance of the workshop to the IR community.

ACKNOWLEDGMENTS