

AKBC 2013: Third Workshop on Automated Knowledge Base Construction

Fabian M. Suchanek
Max Planck Institute for
Informatics
Saarbrücken, Germany
fabian@suchanek.name

Sebastian Riedel
University College London
London, UK
s.riedel@ucl.ac.uk

Sameer Singh
University of Massachusetts
Amherst MA, USA
sameer@cs.umass.edu

Partha P. Talukdar
Carnegie Mellon University
Pittsburgh, USA
ppt@cs.cmu.edu

Abstract

The AKBC 2013 workshop aims to be a venue of excellence and vision in the area of knowledge base construction. This year's workshop will feature keynotes by ten leading researchers in the field, including from Google, Microsoft, Stanford, and CMU. The submissions focus on visionary ideas instead of on experimental evaluation. Nineteen accepted papers will be presented as posters, with nine exceptional papers also highlighted as spotlight talks. Thereby, the workshop aims to provide a vivid forum of discussion about the field of automated knowledge base construction.

Categories and Subject Descriptors

H.2.0 [Information Systems]: Database Management; H.2.8 [Information Systems]: Database Management - Database Applications [Data Mining]; H.3.3 [Information Systems]: Information Storage and Retrieval; I.2.7 [Artificial Intelligence]: Natural Language Processing

General Terms

Algorithms, Performance, Experimentation

Keywords

Information Extraction, Knowledge Bases, Ontologies, Data Mining, Linked Data

1. INTRODUCTION

Topic: Extracting knowledge from Web pages, and integrating it into a coherent knowledge base (KB) is a task that spans the areas of natural language processing, information extraction, information integration, databases, search, and machine learning. Recent years have seen significant advances on the creation of large-scale KBs. Examples include

Wikipedia-based KBs (e.g., YAGO, DBpedia, and Freebase), KBs generated from Web documents (e.g., NELL, PROSPERA), or open information extraction approaches (e.g., TextRunner, PRISMATIC). Further, all major search engines (Yahoo!, Microsoft Bing, and Google) are now being extended using semantic KBs (e.g., Google Knowledge Graph).

The AKBC workshop shall serve as a forum for researchers and practitioners on knowledge base construction. Unlike many other workshops, AKBC puts less emphasis on conventional empirical-based submissions, but more on visionary papers. Its unique characteristic is that it is centered on keynotes by high-profile speakers. This year, the AKBC features ten invited talks from leaders in this area from the academia, industry, and the government. We have invited

- Bonnie Dorr, DARPA, USA
- Evgeniy Gabrilovich, Google Research, USA
- Alon Halevy, Google Research, USA
- Chris Manning, Stanford University, USA
- James Mayfield, Johns Hopkins University, USA
- Andrew McCallum, University of Massachusetts Amherst, USA
- Tom Mitchell, Carnegie Mellon University, USA
- Dan Weld, University of Washington, USA
- Haixun Wang, Microsoft Research Asia

Our objective is to establish AKBC 2013 as a venue of excellence and vision in the area of knowledge base construction. By inviting leading researchers for keynotes, and by focusing particularly on vision paper submissions, we aim to provide a vivid forum of discussion about the field of automated knowledge base construction.

Target Audience: Since our keynote talks are given by prominent researchers in the area (often coordinators of entire scientific projects), the talks are usually high-level and easily understandable. Therefore, we believe that the workshop will be of interest also to novices in the area or first year students who wish to get an overview of automated

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

CIKM'13, Oct. 27–Nov. 1, 2013, San Francisco, CA, USA.

ACM 978-1-4503-2263-8/13/10.

<http://dx.doi.org/10.1145/2505515.2505806>.

KB construction. At the same time, the high calibre of our speakers is almost certain to attract established researchers who wish to get a survey of the latest developments in the field. The vision papers, too, play their role in attracting the audience, as these papers are deliberately designed to provoke thought and discussion from domain experts and novices to the field alike.

With endeavors such as IBM Watson and Google's knowledge graph, knowledge base construction is also an active area for the industry. Since many of the companies that work on semantic data are centered in the Bay Area, we would also actively advertise our workshop to these communities. We hope to arise interest with IBM, Google, Microsoft's Search Labs, and start-up companies.

Past Workshops: This is the third time that the AKBC workshop is being held. The AKBC workshop started in 2010, when Andrew McCallum (University of Massachusetts, Amherst) invited eminent researchers in the area of knowledge extraction to a joint exchange of ideas in Grenoble (France). A dozen keynote talks and eight papers were presented. The papers are available online¹, as are the video recordings of the keynotes². The second AKBC event was held as part of the joint AKBC-WEKEX 2012 workshop held at the NAACL conference. Eleven high profile researchers from industry, academia, and government accepted the invitation to give keynote talks at this two-day workshop. In addition, 23 submissions were accepted as poster papers. The papers are also available online³ as are the recordings of the talks⁴.

2. PAPER SELECTION

The AKBC workshop focuses explicitly on short vision and/or position papers that show not where the research is today, but posit where it would go tomorrow.

Submissions: This year we had 22 submissions from 8 countries, covering topics such as the investigation of reporting bias in knowledge extraction, joint inference and mining of history through the help of knowledge bases.

Reviewing: To improve the quality of our reviews, and to provide the community with the informative discussion that accompanies each paper during the review process, we used an open review paradigm this year. In particular, we used openreview.net for our reviewing process. The papers were submitted to an online platform where both our program committee and members of the general public could provide reviews and comments on a paper, which are all publicly accessible. For each paper we assigned two members of our PC to provide reviews.

The identities of these reviewers was hidden. However, any additional reviews and comments a paper would receive from the public were not anonymous. This was intended to reduce the risk of low-quality comments. Open reviewing has led to high quality reviews and interesting comments that can now be accessed by the community, such as for this paper.

Papers and Best Paper award: Of the 22 submissions we accepted 19 in total. Nine of these 19 papers were accepted as (short) oral presentations. Note that oral papers are also given the opportunity to present a poster. The remaining 10 other accepted papers were chosen as posters only. We will give a best paper award to the most interesting or creative submission, as chosen by the workshop chairs. This year the award goes to

Jonathan Gordon, Benjamin Van Durme, Reporting Bias and Knowledge Acquisition

This paper discusses the mismatch between how often knowledge is reported, and how true it is. This issue is at the heart of knowledge extraction, and important for anyone doing work in this area. The paper not only raises and analyses this issue, but also discusses possible remedies. We believe this paper has the chance to provoke thoughts and shape the field.

3. ACKNOWLEDGMENTS

We would like to convey our thanks to our program committee members.

Doug Downey (Northwestern University), Matt Gardner (Carnegie Mellon University), Rainer Gemulla (Max-Planck Institute for Informatics), Estevam Hruschka (Federal University of Sao Carlos), Jayant Krishnamurthy (Carnegie Mellon University), Zornitsa Kozareva (University of Southern California), Sebastian Michel (Max-Planck Institute for Informatics), Bhavana Dalvi Mishra (Carnegie Mellon University), Marius Pasca (Google Research), Alan Ritter (University of Washington), Ralf Schenkel (Max-Planck Institute for Informatics), Gerhard Weikum (Max-Planck Institute for Informatics), Michael Wick (University of Massachusetts, Amherst), Derry Wijaya (Carnegie Mellon University), Limin Yao (University of Massachusetts, Amherst).

We would like to thank them for their thorough reviews, and also for their readiness to work with an open review system. We would also like to thank the people at openreview.net for their support.

¹<http://akbc.xrce.xerox.com/spip.php?article3>

²http://videlectures.net/akbc2010_grenoble/

³<http://akbcwekex2012.wordpress.com/>

⁴http://videlectures.net/akbcwekex2012_montreal/