NAACL HLT 2019

Extraction of Structured Knowledge from Scientific Publications ESSP

Proceedings of the Workshop

June 6th, 2019
Minneapolis, USA

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-948087-99-5

Introduction

Scientific knowledge is one of the greatest assets of humankind. This knowledge is recorded and disseminated in scientific publications, and the body of scientific literature is growing at an enormous rate. Automatic methods of processing and cataloguing that information are necessary for assisting scientists to navigate this vast amount of information, and for facilitating automated reasoning, discovery and decision making on that data.

Structured information can be extracted at different levels of granularity. Previous and ongoing work has focused on bibliographic information (segmentation and linking of referenced literature), keyword extraction and categorization (e.g., what are tasks, materials and processes central to a publication), and cataloguing research findings. Scientific discoveries can often be represented as pairwise relationships, e.g., protein-protein, drug-drug, and chemical-disease interactions, or as more complicated networks such as action graphs describing scientific procedures (e.g., synthesis recipes in material sciences). Information extracted with such methods can be enriched with time-stamps, and other meta-information, such as indicators of uncertainty or limitations of the discovered facts.

Structured representations, such as knowledge graphs, summarize information from a variety of sources in a convenient and machine readable format. Graph representations that link the information of a large body of publications can reveal patterns and lead to the discovery of new information that would not be apparent from the analysis of just one publication, or from extracted isolated pieces of information. This kind of aggregation can lead to new scientific insights and it can also help to detect trends or find experts for a particular scientific area.

While various workshops have focused separately on several aspects – extraction of information from scientific articles, building and using knowledge graphs, the analysis of bibliographical information, graph algorithms for text analysis – the aim of the ESSP workshop is to elicit and stimulate work that targets the extraction and aggregation of structured information, and to ultimately lead to finding novel information and scientific discoveries.

We have received 15 submissions, of which we accepted 10: 5 for oral presentation, 4 as posters and one demo. The topics covered the biomedical domain, mathematics, computer science and general science, with approaches focusing on various aspects of the extraction, learning, and knowledge processing.

To complement the accepted papers, we welcome four invited speakers from industry, state institutions and academia, to provide insights into knowledge requirements and state of the art in specific fields (medicine, social sciences) and contexts:

Michael Cafarella
University of Michigan
Extraction-Intensive Systems for the Social Sciences

Dina Demner-Fushman
National Library of Medicine
Extracting structured knowledge from biomedical publications

Hoifung Poon
Director, Precision Health NLP @ Microsoft
Machine Reading for Precision Medicine

Chris Welty
Google Research
Just when I thought I was out, they pull me back in – The role of KG in AKBC

We thank our authors, speakers and program committee members for helping us assemble an exciting program on this timely topic. We are grateful to our sponsors – BASF SE Ludwigshafen, the Leibniz Science Campus "Empirical Linguistics and Computational Language Modeling" (LiMo), the German Research Foundation (DFG grant RO5127/2-1) – for making such a diverse and speaker-rich program possible.

Vivi Nastase, Benjamin Roth, Laura Dietz, Andrew McCallum

Organizers:

Vivi Nastase, University of Heidelberg
Benjamin Roth, Ludwig Maximilian University of Munich
Laura Dietz, University of New Hampshire
Andrew McCallum, University of Massachusetts Amherst

Program Committee:

Rabah Al-Zaidy, KAUST, Saudi Arabia
Sergio Baranzini, UCSF
Ken Barker, IBM
Chaitan Baru, UCSD
Chandra Bhagavatula, Allen Institute for AI
Volha Bryl, Springer Nature
Trevor Cohen, MBChB
Anette Frank, University of Heidelberg
Ingo Frommholz, University of Bedfordshire
Daniel Garijo, ISI
Hannaneh Hajishirzi, University of Washington
Keith Hall, Google
Marcel Karnstedt Hulpus, Springer Semantic Web
Bhushan Kotnis, NEC Labs
Anne Lauscher, Mannheim University
Yi Luan, University of Washington
Sebastian Martschat, BASF
Philipp Mayr-Schlegel, GESIS
Arunav Mishra, BASF
Mathias Niepert, NEC Labs
Adam Roegiest, Kira Systems
Martin Schmitt, LMU Munich
Isabel Segura-Bedmar, University Carlos III of Madrid
Mihai Surdeanu, University of Arizona
Niket Tandon, Allen Institute for AI
Karin Verspoor, University of Melbourne
Gerhard Weikum, MPII Saarbruecken
Robert West, EPFL
Guido Zucchon, Queensland University

Invited Speakers:

Michael Cafarella, University of Michigan
Dina Denmer-Fushman, National Library of Medicine
Hoifung Poon, Director, Precision Health NLP @ Microsoft
Chris Welty, Google AI

Table of Contents

Workshop Program

Thursday, June 6, 2019

9:00–10:30 ***Session 1***

9:00–9:15 ***Welcome***

9:15–10:10 *INVITED TALK: Machine Reading for Precision Medicine*
Hoifung Poon

10:10–10:30 *Distantly Supervised Biomedical Knowledge Acquisition via Knowledge Graph Based Attention*
Qin Dai, Naoya Inoue, Paul Reisert, Ryo Takahashi and Kentaro Inui

10:30–11:00 ***Coffee break***

11:00–12:30 ***Session 2***

11:00–11:50 *INVITED TALK: Extraction-Intensive Systems for the Social Sciences*
Michael Cafarella

11:50–12:10 *Scalable, Semi-Supervised Extraction of Structured Information from Scientific Literature*
Kritika Agrawal, Aakash Mittal and Vikram Pudi

12:10–12:30 *Understanding the Polarity of Events in the Biomedical Literature: Deep Learning vs. Linguistically-informed Methods*
Enrique Noriega-Atala, Zhengzhong Liang, John Bachman, Clayton Morrison and Mihai Surdeanu

12:30–14:00 ***Lunch break***

14:00–15:15 ***Session 3***

14:00–14:50 *INVITED TALK: Extracting Structured Knowledge from Biomedical Publications*
Dina Demner-Fushman

14:50–14:55 *Dataset Mention Extraction and Classification*
Animesh Prasad, Chenglei Si and Min-Yen Kan

Thursday, June 6, 2019 (continued)

14:55–15:00 *Annotating with Pros and Cons of Technologies in Computer Science Papers*
Hono Shirai, Naoya Inoue, Jun Suzuki and Kentaro Inui

15:00–15:05 *Browsing Health: Information Extraction to Support New Interfaces for Accessing Medical Evidence*
Soham Parikh, Elizabeth Conrad, Oshin Agarwal, Iain Marshall, Byron Wallace and Ani Nenkova

15:05–15:10 *An Analysis of Deep Contextual Word Embeddings and Neural Architectures for Toponym Mention Detection in Scientific Publications*
Matthew Magnusson and Laura Dietz

15:10–15:15 *STAC: Science Toolkit Based on Chinese Idiom Knowledge Graph*
Meiling Wang, Min Xiao, Changliang Li, Yu Guo, Zhixin Zhao and Xiaonan Liu

15:15–16:00 ***Coffee break and Poster session***

16:00–17:30 ***Session 4***

16:00–16:50 *INVITED TALK: Just When I Thought I Was Out, They Pull Me Back In: The Role of Knowledge Representation in Automatic Knowledge Base Construction*
Chris Welty

16:50–17:10 *Playing by the Book: An Interactive Game Approach for Action Graph Extraction from Text*
Ronen Tamari, Hiroyuki Shindo, Dafna Shahaf and Yuji Matsumoto

17:10–17:30 *Textual and Visual Characteristics of Mathematical Expressions in Scholar Documents*
Vidas Daudaravicius