WOSP 2020

Proceedings of the 8th International Workshop on Mining Scientific Publications

05 August, 2020
Wuhan,
China

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 978-1-952148-88-0

Introduction

The entire body of research literature is currently estimated at 100-150 million publications, with an annual increase of around 1.5 million. Research literature constitutes the complete representation of knowledge we have assembled as human species. It enables us to develop cures to diseases, solve challenging engineering problems and answer many of the world's challenges we are facing today. Systematically reading and analysing the full body of knowledge is now beyond the capacities of any human being. Consequently, it is essential to understand better how we can leverage Natural Language Processing/Text Mining techniques to aid knowledge creation and improve the process by which research is being done.

This workshop aims to bring together people from different backgrounds who:

1. have experience with analysing and mining databases of scientific publications,
2. develop systems that enable such analysis and mining of scientific databases or
3. who develop novel technologies that improve the way research is being done.

The topics of the workshop were organised around the following themes:

1. The whole ecosystem of infrastructures including repositories, aggregators, text-and data-mining facilities, impact monitoring tools, datasets, services and APIs that enable analysis of large volumes of scientific publications.
2. Semantic enrichment of scientific publications utilising text and data mining.
3. analysis of large databases of scientific publications to identify research trends and improve access to research content.

This year, we hosted a new shared task:

3C Citation Context Classification Shared Task

Recent years have witnessed a massive increase in the amount of scientific literature and research data being published online, providing revelation about the advancements in the field of different domains. The introduction of aggregator services like CORE has enabled unprecedented levels of open access to scholarly publications. The availability of full text of the research documents facilitates the possibility of extending the bibliometric studies by identifying the context of the citations. The shared task organized as part of the WOSP 2020 focused on classifying citation context in research publications based on their influence and purpose.

- **Subtask A:** Multiclass classification of citations into one of six classes: Background, Uses, Compare_Contrast, Motivation, Extension and Future.
- **Subtask B:** Binary classification of citations based on the classes Incidental and Influential, a task for identifying the importance of a citation.

Given a citation context, the participants were required to predict the intent of the citations. The participants were provided with a labelled dataset of 3000 training instances annotated using the ACT platform.

We are grateful to the program committee for their careful and thoughtful reviews of the submitted papers. Likewise, we are thankful to the keynote speakers for sharing their research and their vision for the field, and to the workshop attendees for a lively and productive discussion.

Petr Knoth
Christopher Stahl
Bikash Gyawali
David Pride
Suchetha N. Kunnath
Drahomira Herrmannova

Committees

Organisers:

Petr Knoth, Knowledge Media institute, The Open University, UK
Christopher Stahl, Oak Ridge National Laboratory, USA
Bikash Gyawali, Knowledge Media institute, The Open University, UK
David Pride, Knowledge Media institute, The Open University, UK
Suchetha N. Kunnath, Knowledge Media institute, The Open University, UK
Drahomira Herrmannova, Oak Ridge National Laboratory, USA

Program Committee:

Akiko Aizawa, National Insutitute of Informatics, Japan
Iana Atanassova , Université de Bourgogne Franche-Comté, France
Marc Bertin, Université Claude Bernard Lyon 1, France
José Borbinha, Universidade de Lisboa, Portugal
Pravallika Devineni, Oak Ridge National Laboratory, USA
Tirthankar Ghosal, Indian Institute of Technology Patna, India
Saeed-Ul Hassan, Information Technology University, Pakistan
Radim Hladik, Czech Academy of Sciences, Czech Republic
Monica Ihli, University of Tennessee, USA
Roman Kern, Graz University of Technology, Austria
Martin Klein, Los Alamos National Laboratory, USA
Birger Larsen, Aalborg University, Denmark
Paolo Manghi, ISTI-CNR, Italy
Sepideh Mesbah, Delft University of Technology, Netherlands
Peter Mutschke, GESIS Leibniz Institute for the Social Sciences, Germany
Federico Nanni, Alan Turing Institute, UK
Francesco Osborne, The Open University, UK
Robert M. Patton, Oak Ridge National Laboratory, USA
Eloy Rodrigues, Universidade do Minho, Portugal
Wojtek Sylwestrzak, ICM Univeristy of Warsaw, Poland
Vetle Torvik, University of Illinois, USA
Jian Wu, Old Dominion University, USA

Invited Speakers:

Anne Lauscher, University of Mannheim
Allan Hanbury, Vienna University of Technology
David Jurgens, University of Michigan
Neil Smalheiser, University of Illinois at Chicago
Kuansang Wang, MSR Outreach Academic Services

Table of Contents