ACL 2016

The 54th Annual Meeting of the Association for Computational Linguistics

Proceedings of the Student Research Workshop

August 7-12, 2016
Berlin, Germany

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 978-1-945626-02-9

Introduction

Welcome to the ACL 2016 Student Research Workshop!

Following the tradition of the previous years' workshops, we have two tracks: research papers and thesis proposals. The research papers track is as a venue for Ph.D. students, masters students, and advanced undergraduates to describe completed work or work-in-progress along with preliminary results. The thesis proposal track is offered for advanced Ph.D. students who have decided on a thesis topic and are interested in feedback about their proposal and ideas about future directions for their work.

We received in total 60 submissions: 14 thesis proposals and 46 research papers, more than twice as many as were submitted last year. Of these, we accepted 4 thesis proposals and 18 research papers, giving an acceptance rate of 36% overall. This year, all of the accepted papers will be presented as posters alongside the main conference short paper posters on the second day of the conference.

Mentoring programs are a central part of the SRW. This year, students had the opportunity to participate in a pre-submission mentoring program prior to the submission deadline. The mentoring offers students a chance to receive comments from an experienced researcher in the field, in order to improve the quality of the writing and presentation before making their final submission. Nineteen authors participated in the pre-submission mentoring. In addition, authors of accepted papers are matched with mentors who will meet with the students in person during the workshop. This year, each accepted paper is assigned one mentor. Each mentor will prepare in-depth comments and questions prior to the student's presentation, and will provide discussion and feedback during the workshop.

We are very grateful for the generous financial support from Google and the Don and Betty Walker Scholarship Fund. The support of our sponsors allows the SRW to cover the travel and lodging expenses of the authors, keeping the workshop accessible to all students.

We would also like to thank our program committee members for their constructive reviews for each paper and all of our mentors for donating their time to work one-on-one with our student authors. Thank you to our faculty advisers for their advice and guidance, and to the ACL 2016 organizing committee for their constant support. Finally, a huge thank you to all students for their submissions and their participation in this year's SRW. Looking forward to a wonderful workshop!

Organizers:

He He, University of Maryland
Tao Lei, Massachusetts Institute of Technology
Will Roberts, Humboldt-Universität zu Berlin

Faculty Advisors:

Chris Biemann, Technische Universität Darmstadt
Gosse Bouma, Rijksuniversiteit Groningen
Yang Liu, Tsinghua University

Program Committee:

Gabor Angeli, Stanford University
Tania Avgustinova, Saarland University
António Branco, University of Lisbon
Chris Brew, Ohio State University
Wanxiang Che, Harbin Institute of Technology
Danqi Chen, Stanford University
Ann Copestake, University of Cambridge
Michael Elhadad, Ben-Gurion University of the Negev
George Foster, Google
Kevin Gimpel, Toyota Technological Institute at Chicago
Jiang Guo, Harbin Institute of Technology
Nizar Habash, New York University Abu Dhabi
Dilek Hakkani-Tur, Microsoft Research
Keith Hall, Google
Sanda Harabagiu, University of Texas at Dallas
Xiaodong He, Microsoft Research
Lars Hellan, Norwegian University of Science and Technology
Eduard Hovy, Carnegie Mellon University
Philipp Koehn, Johns Hopkins University
Eric Laporte, Université Paris-Est Marne-la-Vallée
Jiwei Li, Stanford University
Junyi Jessy Li, University of Pennsylvania
Wang Ling, Google Deepmind
Wei Lu, Singapore University of Technology and Design
Daniel Marcu, University of Southern California
Taesun Moon, IBM Research
Alessandro Moschitti, Qatar Computing Research Institute
Karthik Narasimhan, Massachusetts Institute of Technology
Graham Neubig, Nara Institute of Science and Technology
Vincent Ng, University of Texas at Dallas
Joakim Nivre, Uppsala University
Kemal Oflazer, Carnegie Mellon University Qatar
Miles Osborne, Bloomberg
Yannick Parmentier, University of Orléans
Nanyun Peng, Johns Hopkins University

Gerald Penn, University of Toronto
Emily Pitler, Google
James Pustejovsky, Brandeis University
Michael Riley, Google
Michael Roth, University of Edinburgh
David Schlangen, Bielefeld University
Satoshi Sekine, New York University
Richard Sproat, Google
Keh-Yih Su, Academia Sinica
Chenhao Tan, Cornell University
Christoph Teichmann, University of Potsdam
Simone Teufel, University of Cambridge
William Yang Wang, Carnegie Mellon University
Bonnie Webber, University of Edinburgh
Bishan Yang, Carnegie Mellon University
Dani Yogatama, Baidu USA
Alessandra Zarcone, Saarland University
Meishan Zhang, Singapore University of Technology and Design
Yuan Zhang, Massachusetts Institute of Technology
Yue Zhang, Singapore University of Technology and Design

Table of Contents

Conference Program

Controlled and Balanced Dataset for Japanese Lexical Simplification
Tomonori Kodaira, Tomoyuki Kajiwara and Mamoru Komachi

Dependency Forest based Word Alignment
Hitoshi Otsuki, Chenhui Chu, Toshiaki Nakazawa and Sadao Kurohashi

Identifying Potential Adverse Drug Events in Tweets Using Bootstrapped Lexicons
Eric Benzschawel

Generating Natural Language Descriptions for Semantic Representations of Human Brain Activity
Eri Matsuo, Ichiro Kobayashi, Shinji Nishimoto, Satoshi Nishida and Hideki Asoh

Improving Twitter Community Detection through Contextual Sentiment Analysis
Alron Jan Lam

Significance of an Accurate Sandhi-Splitter in Shallow Parsing of Dravidian Languages
Devadath V V and Dipti Misra Sharma

Improving Topic Model Clustering of Newspaper Comments for Summarisation
Clare Llewellyn, Claire Grover and Jon Oberlander

Arabizi Identification in Twitter Data
Taha Tobaili

Robust Co-occurrence Quantification for Lexical Distributional Semantics
Dmitrijs Milajevs, Mehrnoosh Sadrzadeh and Matthew Purver

Singleton Detection using Word Embeddings and Neural Networks
Hessel Haagsma

A Dataset for Joint Noun-Noun Compound Bracketing and Interpretation
Murhaf Fares

An Investigation on The Effectiveness of Employing Topic Modeling Techniques to Provide Topic Awareness For Conversational Agents
Omid Moradiannasab

Improving Dependency Parsing Using Sentence Clause Charts
Vincent Kríž and Barbora Hladka

No Day Set (continued)

Graph- and surface-level sentence chunking
Ewa Muszyńska

From Extractive to Abstractive Summarization: A Journey
Parth Mehta

Putting Sarcasm Detection into Context: The Effects of Class Imbalance and Manual Labelling on Supervised Machine Classification of Twitter Conversations
Gavin Abercrombie and Dirk Hovy

Unsupervised Authorial Clustering Based on Syntactic Structure
Alon Daks and Aidan Clark

Suggestion Mining from Opinionated Text
Sapna Negi

An Efficient Cross-lingual Model for Sentence Classification Using Convolutional Neural Network
Yandi Xia, Zhongyu Wei and Yang Liu

QA-It: Classifying Non-Referential It for Question Answer Pairs
Timothy Lee, Alex Lutz and Jinho D. Choi

Building a Corpus for Japanese Wikification with Fine-Grained Entity Classes
Davaajav Jargalsaikhan, Naoaki Okazaki, Koji Matsuda and Kentaro Inui

A Personalized Markov Clustering and Deep Learning Approach for Arabic Text Categorization
Vasu Jindal