NAACL HLT 2016

The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies

Proceedings of the Student Research Workshop

June 12-17, 2016
San Diego, California, USA

©2016 The Association for Computational Linguistics

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 978-1-941643-81-5

Introduction

Welcome to the NAACL-HLT 2016 Student Research Workshop.

This year, we have three different kinds of papers: research papers, thesis proposals, and undergraduate research papers. Thesis proposals were intended for advanced students who have decided on a thesis topic and wish to get feedback on their proposal and broader ideas for their continuing work, while research papers describe completed work or work in progress with preliminary results. In order to encourage undergraduate research, we offered a special track for research papers where the first author is an undergraduate student.

We received 11 research papers, 5 thesis proposals, and 8 undergraduate research papers – making the total number of submissions 24. We accepted 9 research papers, 2 thesis proposals, and 6 undergraduate research papers (17 accepted in total). This translates to an acceptance rate of 81% for research papers, 40% for thesis proposals, and 75% for undergraduate research papers (70% overall). This year, all the SRW papers will be presented at the main conference evening poster session. In addition, each SRW paper is assigned a dedicated mentor. The mentor is an experienced researcher from academia or industry who will prepare in-depth comments and questions in advance for the poster session and will provide feedback to the student author. Thanks to our funding sources, this year's SRW covers registration expenses and provides partial travel and/or lodging support to all student first authors of the SRW papers. We gratefully acknowledge the support from the NSF and Google. We thank our dedicated program committee members who gave constructive and detailed feedback for the student papers. We also would like to thank the NAACL-HLT 2016 organizers and local arrangement chairs.

Organizers:

Jacob Andreas, University of California Berkeley
Eunsol Choi, University of Washington
Angeliki Lazaridou, University of Trento

Faculty Advisors:

Jacob Eisenstein, Georgia Institute of Technology
Nianwen Xue, Brandeis University

Program Committee:

Amjad Abu-Jbara, Microsoft
Bharat Ram Ambati, Apple
Gabor Angeli, Stanford University
Yoav Artzi, Cornell University
Daniel Bauer, Columbia University
Yevgeni Berzak, Massachusetts Institute of Technology
Arianna Bisazza, University of Amsterdam
Shu Cai, University of Southern California
Hiram Calvo, Instituto Politécnico Nacional, Mexico
Dallas Card, Carneige Mellon University
Asli Celikyilmaz, Microsoft
Danqi Chen, Stanford University
Jesse Dodge, University of Washington
Raquel Fernandez, University of Amsterdam
Thomas Francois, UC Louvain
Lea Frermann, University of Edinburgh
Daniel Fried, University of California Berkeley
Annemarie Friedrich, Saarland University
Yoav Goldberg, Bar Ilan University
Amit Goyal , Yahoo! Labs
Michael Hahn, University of Edinburgh
David Hall, Semantic Machines
Eva Hasler, University of Cambridge
Luheng He, University of Washington
John Henderson, MITRE Corporation
Derrick Higgins, Civis Analytics
Dirk Hovy, University of Copenhagen.
Yuening Hu, Yahoo! Labs
Philipp Koehn, University of Edinburgh
Lingpeng Kong, Carneige Mellon University
Ioannis Konstas, University of Washington

Jonathan K. Kummerfeld, University of Berkeley
Kenton Lee, University of Washington
Tal Linzen, Ecole Normale Supérieure
Fei Liu, University of Central Florida
Adam Lopez, University of Edinburgh
Nitin Madnani, ETS
Shervin Malmasi , Harvard University
Diego Marcheggiani, University of Amsterdam
Karthik Narasimhan, Massachusetts Institute of Technology
Arvind Neelakantan, University of Massachusetts Amherst
Denis Paperno, University of Trento
Adam Pauls, University of California, Berkele
Ted Pedersen, University of Minnesota Duluth
Barbara Plank, University of Groningen
Christopher Potts, Stanford University
Daniel Preoţiuc-Pietro, University of Pennsylvania
Preethi Raghavan, IBM T.J. Watson Research Center
Roi Reichart, Technion
Tim Rocktäschel, University College London
Roy Schwartz, The Hebrew University
Minjoon Seo, University of Washington
Kairit Sirts, Tallinn University of Technology
Huan Sun, University of Washington
Swabha Swayamdipta, Carneige Mellon University
Kapil Thadani, Columbia University
Travis Wolfe, Johns Hopkins University
Luke Zettlemoyer, University of Washington

Table of Contents

An End-to-end Approach to Learning Semantic Frames with Feedforward Neural Network
Yukun Feng, Yipei Xu and Dong Yu 1

Analogy-based detection of morphological and semantic relations with word embeddings: what works and what doesn't.
Anna Gladkova, Aleksandr Drozd and Satoshi Matsuoka 8

Argument Identification in Chinese Editorials
Marisa Chow 16

Automatic tagging and retrieval of E-Commerce products based on visual features
Vasu Sharma and Harish Karnick 22

Combining syntactic patterns and Wikipedia's hierarchy of hyperlinks to extract meronym relations
Debela Tesfaye Gemechu, Michael Zock and Solomon Teferra 29

Data-driven Paraphrasing and Stylistic Harmonization
Gerold Hintz 37

Detecting "Smart" Spammers on Social Network: A Topic Model Approach
Linqing Liu, Yao Lu, Ye Luo, Renxian Zhang, Laurent Itti and Jianwei Lu 45

Developing language technology tools and resources for a resource-poor language: Sindhi
Raveesh Motlani 51

Effects of Communicative Pressures on Novice L2 Learners' Use of Optional Formal Devices
Yoav Binoun, Francesca Delogu, Clayton Greenberg, Mindaugas Mozuraitis and Matthew Crocker 59

Explicit Argument Identification for Discourse Parsing In Hindi: A Hybrid Pipeline
Rohit Jain and Dipti Sharma 66

Exploring Fine-Grained Emotion Detection in Tweets
Jasy Suet Yan Liew and Howard R. Turtle 73

Extraction of Bilingual Technical Terms for Chinese-Japanese Patent Translation
Wei Yang, Jinghui Yan and Yves Lepage 81

Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter
Zeerak Waseem and Dirk Hovy 88

Non-decreasing Sub-modular Function for Comprehensible Summarization
Litton J Kurisinkel, Pruthwik Mishra, Vigneshwaran Muralidaran, Vasudeva Varma and Dipti Misra Sharma 94

Phylogenetic simulations over constraint-based grammar formalisms
Andrew Lamont and Jonathan Washington 102

Question Answering over Knowledge Base using Factual Memory Networks
Sarthak Jain 109

Using Related Languages to Enhance Statistical Language Models
Anna Currey, Alina Karakanta and Jon Dehdari 116

Conference Program

Monday June 13th

An End-to-end Approach to Learning Semantic Frames with Feedforward Neural Network
Yukun Feng, Yipei Xu and Dong Yu

Analogy-based detection of morphological and semantic relations with word embeddings: what works and what doesn't.
Anna Gladkova, Aleksandr Drozd and Satoshi Matsuoka

Argument Identification in Chinese Editorials
Marisa Chow

Automatic tagging and retrieval of E-Commerce products based on visual features
Vasu Sharma and Harish Karnick

Combining syntactic patterns and Wikipedia's hierarchy of hyperlinks to extract meronym relations
Debela Tesfaye Gemechu, Michael Zock and Solomon Teferra

Data-driven Paraphrasing and Stylistic Harmonization
Gerold Hintz

Detecting "Smart" Spammers on Social Network: A Topic Model Approach
Linqing Liu, Yao Lu, Ye Luo, Renxian Zhang, Laurent Itti and Jianwei Lu

Developing language technology tools and resources for a resource-poor language: Sindhi
Raveesh Motlani

Tuesday June 14th

Effects of Communicative Pressures on Novice L2 Learners' Use of Optional Formal Devices
Yoav Binoun, Francesca Delogu, Clayton Greenberg, Mindaugas Mozuraitis and Matthew Crocker

Explicit Argument Identification for Discourse Parsing In Hindi: A Hybrid Pipeline
Rohit Jain and Dipti Sharma

Exploring Fine-Grained Emotion Detection in Tweets
Jasy Suet Yan Liew and Howard R. Turtle

Extraction of Bilingual Technical Terms for Chinese-Japanese Patent Translation
Wei Yang, Jinghui Yan and Yves Lepage

Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter
Zeerak Waseem and Dirk Hovy

Non-decreasing Sub-modular Function for Comprehensible Summarization
Litton J Kurisinkel, Pruthwik Mishra, Vigneshwaran Muralidaran, Vasudeva Varma and Dipti Misra Sharma

Phylogenetic simulations over constraint-based grammar formalisms
Andrew Lamont and Jonathan Washington

Question Answering over Knowledge Base using Factual Memory Networks
Sarthak Jain

Using Related Languages to Enhance Statistical Language Models
Anna Currey, Alina Karakanta and Jon Dehdari