

NAACL-HLT 2015 Student Research Workshop (SRW)

**The 2015 Student Research Workshop (SRW)
at the Conference of the North American Chapter
of the Association for Computational Linguistics –
Human Language Technologies**

Proceedings of the Workshop

June 1, 2015
Denver, Colorado, USA

©2015 The Association for Computational Linguistics

Order print-on-demand copies from:

Curran Associates
57 Morehouse Lane
Red Hook, New York 12571
USA
Tel: +1-845-758-0400
Fax: +1-845-758-2633
curran@proceedings.com

ISBN 978-1-941643-50-1

Introduction

Welcome to the NAACL-HLT 2015 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 a record number of submissions this year – 36 research papers, 16 thesis proposals, and 7 undergraduate research papers – making the total number of submissions 59. Out of these, we accepted 9 research papers, 11 thesis proposals, and 3 undergraduate research papers (23 accepted in total). This translates to an acceptance rate of 25% for research papers, 69% for thesis proposals, and 43% for undergraduate research papers (39% overall).

This year, all the SRW papers will be presented at the main conference 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 authors of the SRW papers. We gratefully acknowledge the support from the NSF, Google, Baobab and Fusemachines.

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 2015 organizers and local arrangement chairs – Rada Mihalcea, Joyce Chai, Anoop Sarkar, Priscilla Rasmussen, Matt Post, Adam Lopez, Annie Louis, Kevin B. Cohen, Saif M. Mohammad and Peter Ljunglöf.

Organizers

Student Chairs:

Shibamouli Lahiri, University of Michigan
Karen Mazidi, University of North Texas
Alisa Zhila, Instituto Politécnico Nacional

Faculty Advisors:

Diana Inkpen, University of Ottawa
Smaranda Muresan, Columbia University

Program Committee:

Amjad Abu-Jbara, Microsoft
Gabor Angeli, Stanford University
Yoav Artzi, University of Washington
Beata Beigman Klebanov, Educational Testing Service
Chris Biemann, TU Darmstadt
Arianna Bisazza, University of Amsterdam
Yonatan Bisk, University of Illinois Urbana-Champaign
Jordan Boyd-Graber, University of Colorado
Shu Cai, University of Southern California
Hiram Calvo, Instituto Politécnico Nacional
Asli Celikyilmaz, Microsoft
Monojit Choudhury, Microsoft Research India
Trevor Cohn, University of Melbourne
Hal Daumé III, University of Maryland
Leon Derczynski, University of Sheffield
Kevin Duh, Nara Institute of Science and Technology
Jacob Eisenstein, Georgia Institute of Technology
Aciel Eshky, University of Edinburgh
Kilian Evang, University of Groningen
Paul Felt, Brigham Young University
Michael Flor, Educational Testing Service
Thomas François, UC Louvain
Annemarie Friedrich, Saarland University
Michael Gamon, Microsoft
Qin Gao, Microsoft
Alexander Gelbukh, Instituto Politécnico Nacional
Debanjan Ghosh, Rutgers University
Amit Goyal, Yahoo Labs
Liane Guillou, University of Edinburgh
Eva Hasler, University of Edinburgh

John Henderson, MITRE Corporation
Derrick Higgins, Civis Analytics
Yuening Hu, Yahoo
Ruihong Huang, Stanford University
Héctor Jiménez-Salazar, Universidad Autónoma Metropolitana
Philipp Koehn, University of Edinburgh
Varada Kolhatkar, University of Toronto
Jonathan Kummerfeld, University of California Berkeley
Angeliki Lazaridou, University of Trento
Fei Liu, Carnegie Mellon University
Yang Liu, University of Texas Dallas
Nitin Madnani, Educational Testing Service
Mitch Marcus, University of Pennsylvania
Nisarga Markandaiah, IBM Watson
Thomas Meyer, Google Zurich
Courtney Napoles, Johns Hopkins University
Martha Palmer, University of Colorado
Ted Pedersen, University of Minnesota Duluth
Christopher Potts, Stanford University
Vinodkumar Prabhakaran, Columbia University
Rashmi Prasad, University of Wisconsin Milwaukee
Preethi Raghavan, IBM TJ Watson Research Center, Yorktown Heights NY
Owen Rambow, Columbia University
Sravana Reddy, Dartmouth College
Roi Reichart, Technion
Philip Resnik, University of Maryland
Eduardo Rodriguez, Instituto Politécnico Nacional
Kairit Sirts, Tallinn University of Technology
Thamar Solorio, University of Houston
Swapna Somasundaran, Educational Testing Service
Kapil Thadani, Columbia University
Eva Maria Vecchi, University of Cambridge
Nina Wacholder, Rutgers University
Jason Williams, Microsoft
Travis Wolfe, Johns Hopkins University
Xuchen Yao, Johns Hopkins University
Luke Zettlemoyer, University of Washington
Qiuye Zhao, University of Pennsylvania

Table of Contents

<i>Cache-Augmented Latent Topic Language Models for Speech Retrieval</i>	
Jonathan Wintrode	1
<i>Reliable Lexical Simplification for Non-Native Speakers</i>	
Gustavo Paetzold	9
<i>Analyzing Newspaper Crime Reports for Identification of Safe Transit Paths</i>	
Vasu Sharma, Rajat Kulshreshtha, Puneet Singh, Nishant Agrawal and Akshay Kumar	17
<i>Relation extraction pattern ranking using word similarity</i>	
Konstantinos Lambrou-Latreille	25
<i>Towards a Better Semantic Role Labeling of Complex Predicates</i>	
Glorianna Jagfeld and Lonneke van der Plas	33
<i>Exploring Relational Features and Learning under Distant Supervision for Information Extraction Tasks</i>	
Ajay Nagesh	40
<i>Entity/Event-Level Sentiment Detection and Inference</i>	
Lingjia Deng	48
<i>Initial Steps for Building a Lexicon of Adjectives with Scalemates</i>	
Bryan Wilkinson	57
<i>A Preliminary Evaluation of the Impact of Syntactic Structure in Semantic Textual Similarity and Semantic Relatedness Tasks</i>	
Ngoc Phuoc An Vo and Octavian Popescu	64
<i>Benchmarking Machine Translated Sentiment Analysis for Arabic Tweets</i>	
Eshrag Refaee and Verena Rieser	71
<i>Learning Kernels for Semantic Clustering: A Deep Approach</i>	
Ignacio Arroyo-Fernández	79
<i>Narrowing the Loop: Integration of Resources and Linguistic Dataset Development with Interactive Machine Learning</i>	
Seid Muhie Yimam	88
<i>Relation Extraction from Community Generated Question-Answer Pairs</i>	
Denis Savenkov, Wei-Lwun Lu, Jeff Dalton and Eugene Agichtein	96
<i>Detecting Translation Direction: A Cross-Domain Study</i>	
Sauleh Eetemadi and Kristina Toutanova	103

<i>Improving the Translation of Discourse Markers for Chinese into English</i>	
David Steele	110
<i>Discourse and Document-level Information for Evaluating Language Output Tasks</i>	
Carolina Scarton	118
<i>Speeding Document Annotation with Topic Models</i>	
Forough Poursabzi-Sangdeh and Jordan Boyd-Graber	126
<i>Lifelong Machine Learning for Topic Modeling and Beyond</i>	
Zhiyuan Chen	133
<i>Semantics-based Graph Approach to Complex Question-Answering</i>	
Tomasz Jurczyk and Jinho D. Choi	140
<i>Recognizing Textual Entailment using Dependency Analysis and Machine Learning</i>	
Nidhi Sharma, Richa Sharma and Kanad K. Biswas	147
<i>Bilingual lexicon extraction for a distant language pair using a small parallel corpus</i>	
Ximena Gutierrez-Vasques	154
<i>Morphological Paradigms: Computational Structure and Unsupervised Learning</i>	
Jackson Lee	161
<i>Computational Exploration to Linguistic Structures of Future: Classification and Categorization</i>	
Aiming Ni, Jinho D. Choi, Jason Shepard and Phillip Wolff	168

Student Research Workshop Program

Monday, June 1, 2015

18:00–21:00 Poster Session

Thesis Proposals

Reliable Lexical Simplification for Non-Native Speakers

Gustavo Paetzold

Relation extraction pattern ranking using word similarity

Konstantinos Lambrou-Latreille

Exploring Relational Features and Learning under Distant Supervision for Information Extraction Tasks

Ajay Nagesh

Entity/Event-Level Sentiment Detection and Inference

Lingjia Deng

Learning Kernels for Semantic Clustering: A Deep Approach

Ignacio Arroyo-Fernández

Narrowing the Loop: Integration of Resources and Linguistic Dataset Development with Interactive Machine Learning

Seid Muhie Yimam

Improving the Translation of Discourse Markers for Chinese into English

David Steele

Discourse and Document-level Information for Evaluating Language Output Tasks

Carolina Scarton

Lifelong Machine Learning for Topic Modeling and Beyond

Zhiyuan Chen

Bilingual lexicon extraction for a distant language pair using a small parallel corpus

Ximena Gutierrez-Vasques

Morphological Paradigms: Computational Structure and Unsupervised Learning

Jackson Lee

Monday, June 1, 2015 (continued)

Graduate Student Research Papers

Cache-Augmented Latent Topic Language Models for Speech Retrieval

Jonathan Wintrobe

Initial Steps for Building a Lexicon of Adjectives with Scalemates

Bryan Wilkinson

A Preliminary Evaluation of the Impact of Syntactic Structure in Semantic Textual Similarity and Semantic Relatedness Tasks

Ngoc Phuoc An Vo and Octavian Popescu

Benchmarking Machine Translated Sentiment Analysis for Arabic Tweets

Eshrag Refaee and Verena Rieser

Relation Extraction from Community Generated Question-Answer Pairs

Denis Savenkov, Wei-Lwun Lu, Jeff Dalton and Eugene Agichtein

Detecting Translation Direction: A Cross-Domain Study

Sauleh Eetemadi and Kristina Toutanova

Speeding Document Annotation with Topic Models

Forough Poursabzi-Sangdeh and Jordan Boyd-Graber

Semantics-based Graph Approach to Complex Question-Answering

Tomasz Jurczyk and Jinho D. Choi

Recognizing Textual Entailment using Dependency Analysis and Machine Learning

Nidhi Sharma, Richa Sharma and Kanad K. Biswas

Undergraduate Student Research Papers

Analyzing Newspaper Crime Reports for Identification of Safe Transit Paths

Vasu Sharma, Rajat Kulshreshtha, Puneet Singh, Nishant Agrawal and Akshay Kumar

Towards a Better Semantic Role Labeling of Complex Predicates

Glorianna Jagfeld and Lonneke van der Plas

Computational Exploration to Linguistic Structures of Future: Classification and Categorization

Aiming Ni, Jinho D. Choi, Jason Shepard and Phillip Wolff