

NAACL HLT SRW 2013

**Proceedings of the NAACL-HLT 2013 Student Research
Workshop**

Proceedings

9–14 June 2013

©2013 The Association for Computational Linguistics

209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-937284-47-3

Introduction

Welcome to the NAACL HLT 2013 Student Research Workshop.

This year, we have two different kinds of paper: research papers and thesis proposals. Thesis proposals are 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 can describe completed work or work in progress with preliminary results.

All the papers will be presented in the main conference poster session, giving the opportunity for students to interact and present their work to a large and diverse audience. In addition, we have a separate session for the student papers on the first day of workshops (after the main conference). During this session, students will present their papers and receive feedback from mentors. The mentors are experienced researchers who will prepare in-depth comments and questions in advance of the presentation. Each accepted paper is assigned a mentor. The separate session is newly introduced this year and differs from recent NAACL student workshops where student talks were during the main conference sessions or the papers were presented as posters only. We expect that the focused workshop will provide a greater opportunity for receive feedback from mentors, and also allow the students to network and socialize with other student participants.

We received 8 thesis proposals and 15 research papers. Out of these we accepted 6 thesis proposals and 7 research papers leading to an acceptance rate of 75% for thesis proposals and 47% for research papers. We thank our dedicated program committee who gave constructive and detailed reviews for the student papers. We also thank the NAACL 2013 organizing committee—Lucy Vanderwende, Hal Daumé III, Katrin Kirchhoff, Priscilla Rassmussen, Matt Post and Colin Cherry.

Student Chairs:

Annie Louis, University of Edinburgh
Richard Socher, Stanford University

Faculty Advisors:

Julia Hockenmaier, University of Illinois at Urbana-Champaign
Eric Ringger, Brigham Young University

Program Committee:

Yukino Baba, University of Tokyo
Emily Bender, University of Washington
Jonathan Berant, Stanford University
Chris Biemann, University of Darmstadt
Yonatan Bisk, University of Illinois at Urbana-Champaign
Jackie Chi Kit Cheung, University of Toronto
Mark Dredze, Johns Hopkins University
Kevin Duh, NAIST
Jacob Eisenstein, Georgia Institute of Technology
Jason Eisner, Johns Hopkins University
Paul Felt, Brigham Young University
Jennifer Gillenwater, University of Pennsylvania
David Hall, University of California, Berkeley
Derrick Higgins, Educational Testing Service
Yuening Hu, University of Maryland, College Park
Kevin Knight, University of Southern California
Philip Koehn, University of Edinburgh
Diane Littman, University of Pittsburgh
Fei Liu, Bosch Research
Yang Liu, University of Texas at Dallas
Bill Lund, Brigham Young University
Rebecca Mason, Brown University
Rada Mihalcea, University of North Texas
Christopher Potts, Stanford University
Vahed Qazvinian, Google
Preethi Raghavan, The Ohio State University
Marta Recasens, Stanford University
Sravana Reddy, Dartmouth University
Chenhao Tan, Cornell University
Kapil Thadani, Columbia University
Scott Yih, Microsoft Research
Qiuye Zhao, University of Pennsylvania

Table of Contents

<i>Critical Reflections on Evaluation Practices in Coreference Resolution</i>	
Gordana Ilic Holen	1
<i>Reducing Annotation Effort on Unbalanced Corpus based on Cost Matrix</i>	
Wencan Luo, Diane Litman and Joel Chan	8
<i>A Machine Learning Approach to Automatic Term Extraction using a Rich Feature Set</i>	
Merley Conrado, Thiago Pardo and Solange Rezende	16
<i>A Rule-based Approach for Karmina Generation</i>	
Franky Franky	24
<i>From Language to Family and Back: Native Language and Language Family Identification from English Text</i>	
Ariel Stolerman, Aylin Caliskan and Rachel Greenstadt	32
<i>Ontology Label Translation</i>	
Mihael Arcan and Paul Buitelaar	40
<i>Reversing Morphological Tokenization in English-to-Arabic SMT</i>	
Mohammad Salameh, Colin Cherry and Grzegorz Kondrak	47
<i>Statistical Machine Translation in Low Resource Settings</i>	
Ann Irvine	54
<i>Large-Scale Paraphrasing for Natural Language Understanding</i>	
Juri Ganitkevitch	62
<i>Domain-Independent Captioning of Domain-Specific Images</i>	
Rebecca Mason	69
<i>Helpfulness-Guided Review Summarization</i>	
Wenting Xiong	77
<i>Entrainment in Spoken Dialogue Systems: Adopting, Predicting and Influencing User Behavior</i>	
Rivka Levitan	84
<i>User Goal Change Model for Spoken Dialog State Tracking</i>	
Yi Ma	91

Workshop Program

Thursday, June 13, 2013

9:00–9:15 Opening remarks

Session 1: Research paper presentations

9:15–9:30 *Critical Reflections on Evaluation Practices in Coreference Resolution*
Gordana Ilic Holen

9:30–9:45 *Reducing Annotation Effort on Unbalanced Corpus based on Cost Matrix*
Wencan Luo, Diane Litman and Joel Chan

9:45–10:00 *A Machine Learning Approach to Automatic Term Extraction using a Rich Feature Set*
Merley Conrado, Thiago Pardo and Solange Rezende

10:00–10:15 *A Rule-based Approach for Karmina Generation*
Franky Franky

10:15–10:30 *From Language to Family and Back: Native Language and Language Family Identification from English Text*
Ariel Stolerman, Aylin Caliskan and Rachel Greenstadt

10:30–11:00 Coffee break

Session 2: Research paper presentations

11:00–11:15 *Ontology Label Translation*
Mihael Arcan and Paul Buitelaar

11:15–11:30 *Reversing Morphological Tokenization in English-to-Arabic SMT*
Mohammad Salameh, Colin Cherry and Grzegorz Kondrak

Thursday, June 13, 2013 (continued)

Session 3: Thesis proposal presentations

11:30–12:00 *Statistical Machine Translation in Low Resource Settings*
Ann Irvine

12:00–12:30 *Large-Scale Paraphrasing for Natural Language Understanding*
Juri Ganitkevitch

12:30–14:00 Lunch

Session 4: Thesis proposal presentations

14:00–14:30 *Domain-Independent Captioning of Domain-Specific Images*
Rebecca Mason

14:30–15:00 *Helpfulness-Guided Review Summarization*
Wenting Xiong

15:00–15:30 *Entrainment in Spoken Dialogue Systems: Adopting, Predicting and Influencing User Behavior*
Rivka Levitan

15:30–16:00 Coffee break

Session 5: Thesis proposal presentation

16:00–16:30 *User Goal Change Model for Spoken Dialog State Tracking*
Yi Ma

16:30–17:30 Panel