

ACL-IJCNLP 2009

**Joint Conference of the
47th Annual Meeting of the
Association for Computational Linguistics
and
4th International Joint Conference on
Natural Language Processing
of the AFNLP**

Proceedings of the Student Research Workshop

4 August 2009
Suntec, Singapore

Production and Manufacturing by
World Scientific Publishing Co Pte Ltd
5 Toh Tuck Link
Singapore 596224

©2009 The Association for Computational Linguistics
and The Asian Federation of Natural Language Processing

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

Introduction

Welcome to the ACL-IJCNLP 2009 Student Research Workshop! The Student Research Workshop is now an established tradition at ACL conferences and provides a venue for student researchers investigating topics in Computational Linguistics and Natural Language Processing to present their work and receive feedback. This year we received a total of 25 submissions coming from 15 different countries, and accepted 12 of them. 5 will be presented orally, and 7 as posters, during a common poster session with the main conference. A total of 43 students and senior researchers agreed to serve on the program committee, which allowed us to assign 3 reviewers per paper. We would like to thank the reviewers for understanding the spirit of the Student Research Workshop and giving careful and constructive reviews. We hope their comments will be helpful to all the students who submitted their work. All presenters were offered travel grants to assist them in their travel to Singapore, thanks to generous support from the U.S. National Science Foundation, The Asian Federation of Natural Language Processing, The Nagao Fund of the AFNLP, and The Walker Fund of the Association for Computational Linguistics.

We are very grateful to Brian Roark and Grace Ngai, our faculty advisors, for their advice, constant support (and reminders!), and obtaining of funding. Finally, we would like to thank the general chair of ACL-IJCNLP 2009, Keh-Yih Su, the program chairs, Jian Su and Janyce Wiebe, the publications chairs Regina Barzilay and Jing-Shin Chang, Haizhou Li and the local organization committee, and Priscilla Rasmussen.

Davis Dimalen, Jenny Rose Finkel, and Blaise Thomson
The ACL-IJCNLP 2009 Student Research Workshop co-chairs

Faculty advisors:

Grace Ngai
Hong Kong Polytechnic University
Kowloon, Hong Kong

Brian Roark
Oregon Health and Science University
Beaverton, Oregon, USA

Chairs:

Davis Muhajereen D. Dimalen
CLCLP, Taiwan International Graduate Program
Academia Sinica, Taiwan

Jenny Rose Finkel
Stanford University
Stanford, California, USA

Blaise Thomson
Cambridge University
Cambridge, UK

Program Committee:

Galen Andrew, Microsoft, USA
Eva Banik, Open University, UK
Shane Bergsma, University of Alberta, Canada
Dan Bohus, Microsoft, USA
Don Erick Bonus, Jose Rizal University, Philippines
Wauter Bosma, Vrije Universiteit, Netherlands
Bill Byrne, University of Cambridge, UK
Colin Cherry, Microsoft, USA
Huang Chu-Ren, Academia Sinica, Taiwan
Shay Cohen, Carnegie Mellon University, USA
Editha D. Dimalen, Mindanao State University, Philippines
Mark Dredze, University of Pennsylvania, USA
Jacob Eisenstein, University of Illinois at Urbana-Champaign, USA
Sharon Goldwater, University of Edinburgh, UK
Mark Greenwood, University of Sheffield, UK
Masato Hagiwara, Nagoya University, Japan
David Hall, Stanford University, USA

LI Haizhou, Institute for Infocomm Research, Singapore
Aurelie Herbelot, University of Cambridge, UK
Samar Husain, IIIT Hyderabad, India
Pei-Yun Hsueh, University of Edinburgh, UK and IBM, USA
Sanaz Jabbari, University of Sheffield, UK
Maggie LI (Li Wenjie Maggie), Hong Kong Polytechnic University, China
Yuji Matsumoto, NAIIST, Japan
David McClosky, Brown University, USA
Roser Morante, University of Antwerp, Belgium
Teruhisa Misu, NICT/ATR, Japan
Vincent Ng, University of Texas at Dallas, USA
Patrick Pantel, Yahoo, USA
Jong C. Park, KAIST, Korea
JIN Peng, Peking University, China
Le Hong Phuong, INRIA Lorraine, France
Emily Pitler, University of Pennsylvania, USA
Daniel Ramage, Stanford University, USA
Antti-Veikko Rosti, BBN, USA
Rachel Edita Roxas, De LaSalle University-Manila, Philippines
Philipp Spanger, Tokyo Institute of Technology, Japan
Reut Tsarfaty, University of Amsterdam, Netherlands
Joseph Turian, University of Montreal, Canada
Lonneke van der Plas, University of Geneva, Switzerland
Sumithra Velupillai, Stockholm University/KTH, Sweden
Andreas Vlachos, University of Cambridge, UK
Liang-Chih Yu, Yuan-Ze University, China
JIA Yuxiang, Peking University, China

Table of Contents

<i>Sense-based Interpretation of Logical Metonymy Using a Statistical Method</i> Ekaterina Shutova	1
<i>Insights into Non-projectivity in Hindi</i> Prashanth Mannem, Himani Chaudhry and Akshar Bharati	10
<i>Annotating and Recognising Named Entities in Clinical Notes</i> Yefeng Wang	18
<i>Paraphrase Recognition Using Machine Learning to Combine Similarity Measures</i> Prodromos Malakasiotis	27
<i>A System for Semantic Analysis of Chemical Compound Names</i> Henriette Engelken	36
<i>Sentence diagram generation using dependency parsing</i> Elijah Mayfield	45
<i>Accurate Learning for Chinese Function Tags from Minimal Features</i> Caixia Yuan, Fuji Ren and Xiaojie Wang	54
<i>Optimizing Language Model Information Retrieval System with Expectation Maximization Algorithm</i> Justin Liang-Te Chiu and Jyun-Wei Huang	63
<i>Data Cleaning for Word Alignment</i> Tsuyoshi Okita	72
<i>The Modulation of Cooperation and Emotion in Dialogue: The REC Corpus</i> Federica Cavicchio	81
<i>Clustering Technique in Multi-Document Personal Name Disambiguation</i> Chen Chen, Hu Junfeng and Wang Houfeng	88
<i>Creating a Gold Standard for Sentence Clustering in Multi-Document Summarization</i> Johanna Geiss	96

Conference Program

Student Research Workshop, 4 August, 2009

Oral Session (10:15–12:20)

- 10:15–10:40 *Sense-based Interpretation of Logical Metonymy Using a Statistical Method*
Ekaterina Shutova
- 10:40–11:05 *Insights into Non-projectivity in Hindi*
Prashanth Mannem, Himani Chaudhry and Akshar Bharati
- 11:05–11:30 *Annotating and Recognising Named Entities in Clinical Notes*
Yefeng Wang
- 11:30–11:55 *Paraphrase Recognition Using Machine Learning to Combine Similarity Measures*
Prodromos Malakasiotis
- 11:55–12:20 *A System for Semantic Analysis of Chemical Compound Names*
Henriette Engelken

Poster Session (12:20–14:20)

Sentence diagram generation using dependency parsing
Elijah Mayfield

Accurate Learning for Chinese Function Tags from Minimal Features
Caixia Yuan, Fuji Ren and Xiaojie Wang

Optimizing Language Model Information Retrieval System with Expectation Maximization Algorithm
Justin Liang-Te Chiu and Jyun-Wei Huang

Data Cleaning for Word Alignment
Tsuyoshi Okita

The Modulation of Cooperation and Emotion in Dialogue: The REC Corpus
Federica Cavicchio

Student Research Workshop, 4 August, 2009 (continued)

Clustering Technique in Multi-Document Personal Name Disambiguation

Chen Chen, Hu Junfeng and Wang Houfeng

Creating a Gold Standard for Sentence Clustering in Multi-Document Summarization

Johanna Geiss