ACL-05

Proceedings of the Student Research Workshop

June 27, 2005
University of Michigan
Ann Arbor, Michigan, USA

Production and Manufacturing by
Omnipress Inc.
Post Office Box 7214
Madison, WI 53707-7214

Sponsored by
The National Science Foundation

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
75 Paterson Street, Suite 9
New Brunswick, NJ 08901
USA
Tel: +1-732-342-9100
Fax: +1-732-342-9339
acl@aclweb.org

Introduction

Welcome to the ACL Student Research Workshop! We had an amazing amount of student participation this year. We received a record number of submissions – over 70 papers from students in 19 countries. In order to accommodate the overwhelming response, we've expanded the Student Research Workshop. In addition to the regular paper presentations it now includes a poster session. The acceptance rate for the regular paper presentations was 11% (more competitive than the main conference!), and the acceptance rate for the poster presentations from the remaining submissions was 28%. These proceedings contain papers for both the regular and the poster presentations.

We are grateful to our faculty advisor, Regina Barzilay, and to Mary Harper of the National Science Foundation for organizing sponsorship from the National Science Foundation for the Student Research Workshop. The NSFs generous grant has paid for the conference registration fees for all of the student workshop participants and a good portion of their travel expenses. We are also grateful for the Don and Betty Walker International Student Fund which has provided assistance for two students who had their work accepted in the main conference.

We would like to extend our thanks to the students and faculty on the program committee who dutifully reviewed the papers and gave useful feedback to everyone who submitted papers, and to all the students who submitted such excellent work.

Enjoy the workshop!

Chris Callison-Burch and Stephen Wan
ACL Student Research Workshop Co-Chairs

Co-Chairs:

Chris Callison-Burch (Univeristy of Edinburgh)
Stephen Wan (Macquarie University)

Faculty Advisor:

Regina Barzilay (MIT)

Program Committee:

Laura Alonso (Barcelona)
Timothy Baldwin (Melbourne)
Colin Bannard (Edinburgh)
Phil Blunsom (Melbourne)
Bernd Bohnet (Stuttgart)
Cem Bozsahin (Middle East Technical University)
Chris Brew (Ohio State)
Marine Carpuat (HKUST)
Stephen Clark (Oxford)
Trevor Cohn (Melbourne)
James Curran (Sydney)
Tiphaine Dalmas (Edinburgh)
Hal Daume III (ISI)
Mona Diab (Stanford)
Mark Dras (Macquarie)
Pablo Duboue (IBM)
Elena Filatova (Columbia)
Erin Fitzgerald (JHU)
Dan Flickinger (Stanford)
Pablo Gamallo (Universidade Nova de Lisboa)
Claire Grover (Edinburgh)
Graeme Hirst (Toronto)
Julia Hockenmaier (U Penn)
Ben Hutchinson (Edinburgh)
Frank Keller (Edinburgh)
Simon King (Edinburgh)
Alistair Knott (Otago)
Philipp Koehn (Edinburgh)
Emiel Krahmer (Tilburg)
Corrin Lakeland (Otago)
Mirella Lapata (Edinburgh)
Alon Lavie (CMU)
Maria Liakata (Oxford)

Daniel Marcu (ISI)
Daniel Midgley (University of Western Australia)
Diego Molla (Macquarie)
Ani Nenkova (Columbia)
Leif Nielsen (King's College)
Bo Pang (Cornell)
Cecile Paris (CSIRO)
Jean-Philippe Prost (Macquarie)
Steve Renals (Edinburgh)
Philip Resnik (Maryland)
Graeme Ritchie (Aberdeen)
Charles Schafer (JHU)
Advaith Siddharthan (Columbia)
Michel Simard (Xerox)
Matthew Stone (Rutgers)
Mark Swerts (Tilburg)
David Talbot (Edinburgh)
Leonoor van der Beek (Groningen)
Florian Wolf (Cambridge)

Table of Contents

Student Research Workshop Program

Monday, June 27, 2005: Presentations

Student Presentations: Session 1

9:00–9:30 *Hybrid methods for POS guessing of Chinese unknown words*
Xiaofei Lu

9:30–10:00 *Understanding the thematic structure of the Qur'an: an exploratory multivariate approach*
Naglaa Thabet

10:00–10:30 *An Extensive Empirical Study of Collocation Extraction Methods*
Pavel Pecina

Student Presentations: Session 2

2:30–3:00 *Jointly Labeling Multiple Sequences: A Factorial HMM Approach*
Kevin Duh

3:00–3:30 *Exploiting Named Entity Taggers in a Second Language*
Thamar Solorio

Student Presentations: Session 3

4:00–4:30 *Automatic Discovery of Intentions in Text and its Application to Question Answering*
Marta Tatu

4:30–5:00 *American Sign Language Generation: Multimodal NLG with Multiple Linguistic Channels*
Matt Huenerfauth

5:00–5:30 *Using Emoticons to reduce Dependency in Machine Learning Techniques for Sentiment Classification*
Jonathon Read

Monday, June 27, 2005: Posters

All posters will be on display during the Student Lunch from 12:00–1:30

Learning Meronyms from Biomedical Text
Angus Roberts

Using Readers to Identify Lexical Cohesive Structures in Texts
Beata Beigman Klebanov

Towards an Optimal Lexicalization in a Natural-Sounding Portable Natural Language Generator for Dialog Systems
Inge M. R. De Bleecker

Phrase Linguistic Classification and Generalization for Improving Statistical Machine Translation
Adrià de Gispert

Automatic Induction of a CCG Grammar for Turkish
Ruken Cakici

Dialogue Act Tagging for Instant Messaging Chat Sessions
Edward Ivanovic

Learning Strategies for Open-Domain Natural Language Question Answering
Eugene Grois

Dependency-Based Statistical Machine Translation
Heidi Fox

Minimalist Parsing of Subjects Displaced from Embedded Clauses in Free Word Order Languages
Asad B. Sayeed

Centrality Measures in Text Mining: Prediction of Noun Phrases that Appear in Abstracts
Zhuli Xie

A corpus-based approach to topic in Danish dialog
Philip Diderichsen and Jakob Elming

(posters continued)

Learning Information Structure in The Prague Treebank
Oana Postolache

Speech Recognition of Czech - Inclusion of Rare Words Helps
Petr Podvesky and Pavel Machek

Using Bilingual Dependencies to Align Words in Enlish/French Parallel Corpora
Sylwia Ozdowska

An Unsupervised System for Identifying English Inclusions in German Text
Beatrice Alex

Corpus-Oriented Development of Japanese HPSG Parsers
Kazuhiro Yoshida

Unsupervised Discrimination and Labeling of Ambiguous Names
Anagha Kulkarni

A Domain-Specific Statistical Surface Realizer
Jeffrey Russell