ACL 2007

Proceedings of the Student Research Workshop

June 25–26, 2007

Prague, Czech Republic

Production and Manufacturing by
Omnipress
2600 Anderson Street
Madison, WI 53704
USA

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

Preface

On behalf of the Organizing Committee, we are pleased to present the proceedings of the Student Research Workshop held at the 45th Annual Meeting of the Association for Computational Linguistics (ACL) in Prague, Czech Republic, June 25–27, 2007. The Student Research Workshop is an established tradition at the Annual Meetings of the Association for Computational Linguistics and builds on the success of no less than 16 previous student sessions at ACL meetings.

Students in Computational Linguistics, Natural Language Processing and related fields were offered the possibility to present their work in a setting embedded in the main conference. The workshop plays an integral role in ACL's efforts to build and maintain a research community by investing in young researchers that will shape the field in the years to come. The workshop aimed at providing feedback from senior to beginner researchers. In the call for papers, we explicitly aimed at students in an early stage of their Ph.D. work. We felt that this group could gain the most benefit from this event, as the experts' feedback can still influence their research directions.

The Program Committee was compiled such that about half of the reviewers were students or young researchers, and the other half consisted of senior scientists. This mixture ensures that the scientific quality of reviews is high, while student-specific issues are well understood by the committee members. We are indebted to our 52 reviewers for their elaborate, thoughtful and high quality reviews, which will also be of great help to those students whose work could not be accepted for presentation.

We received 52 submissions from all over the world, of which 16 were accepted for presentation: 9 for oral presentation and 7 for poster presentation. The presentation format was assigned based on thoughts about how the work could be presented best, and does not indicate a quality difference among papers, which are all fixed to the same length of 6 pages.

This year's workshop features contributions from a wide range of topics. Various issues on grammar are dealt with in five papers: Richard Johansson uses logistic online learning for incremental dependency parsing, Nathan C. Sanders measures syntactic differences in British English, Elias Ponvert induces combinatory categorial grammars with genetic algorithms, Bart Cramer investigates limitations of current grammar induction techniques, and Aleksander Buczyński describes an implementation that combines partial parsing and morphosyntactic disambiguation.

Another five contributions can be subsumed under the scope of semantics: Radosław Moszczyński provides a classification of multi-word expressions especially for highly inflected languages, Paul Nulty classifies noun phrases along semantic properties using web counts and machine learning, Diarmuid Ó Séaghdha annotates and learns compound noun semantics, Kata Gábor and Enikő Héja cluster Hungarian verbs by complementation patterns, and Silke Scheible lays out foundations of a computational treatment of superlatives.

Research on dialects and different languages is carried out by three papers: Andrea Mulloni performs cognate prediction in a bilingual setting, Yves Scherrer presents adaptive measures to graphemic similarity for inducing dialect lexicons, and Jelena Prokić identifies linguistic structure in a quantitative analysis of Bulgarian dialects. For opinionated Chinese Information Retrieval, Taras Zagibalov examines the utility of various features. Structuring texts is the topic of two papers: Olena Medelyan uses graph clustering to compute lexical chains, and Martina Naughton exploits structure for event discovery using the MDI algorithm.

Following the workshop tradition, a panel of senior researchers will take part in the presentation of papers, providing in-depth comments on the work of each author either immediately after the oral presentation or in front of the poster. We would like to thank the panelists in advance for fulfilling such an important role.

Many people contributed to the success of this year's Student Research Workshop. Apart from Program Committee members and panelists, we would like to thank the ACL conference organizers for involving us in their planning, the webmasters for swiftly handling update requests for the SRW page, the publication chair for providing us the facilities to compile this volume, and, most of all, the students for their hard work in preparing their submissions. Finally, we are grateful to the National Science Foundation for generously sponsoring our event: All student presenters received reimbursement of registration and accommodation as well as almost full coverage of travel costs.

The ACL 2007 Student Research Workshop Co-Chairs
Chris Biemann, Violeta Seretan, Ellen Riloff

Organizers

Chairs:

Chris Biemann, University of Leipzig, Germany
Violeta Seretan, University of Geneva, Switzerland

Faculty advisor:

Ellen Riloff, University of Utah, USA

Program Committee:

Laura Alonso i Alemany, Universidad de la República, Uruguay
and Universidad Nacional de Córdoba, Argentina
Galia Angelova, Bulgarian Academy of Sciences, Bulgaria
Timothy Baldwin, University of Melbourne, Australia
Raffaella Bernardi, Free University of Bozen-Bolzano, Italy
Stephan Bloehdorn, University of Karlsruhe, Germany
Gemma Boleda, Universitat Pompeu Fabra, Spain
Kalina Bontcheva, University of Sheffield, UK
Monojit Choudhury, Indian Institute of Technology, Kharagpur, India
Philipp Cimiano, University of Karlsruhe, Germany
Alexander Clark, Royal Holloway, University of London, UK
Gaël Harry Dias, University of Beira Interior, Portugal
Katrin Erk, University of Texas at Austin, USA
Stefan Evert, University of Osnabrück, Germany
Afsaneh Fazly, University of Toronto, Canada
Alexander Gelbukh, National Polytechnic Institute, Mexico
Alfio Gliozzo, ITC-irst, Trento, Italy
Yoav Goldberg, Ben-Gurion University of the Negev, Israel
Jean-Philippe Goldman, University of Geneva, Switzerland
Günther Görz, University of Erlangen, Germany
Iryna Gurevych, Darmstadt University of Technology, Germany
Catalina Hallett, The Open University, UK
Laura Hasler, University of Wolverhampton, UK
Janne Bondi Johannessen, University of Oslo, Norway
Philipp Koehn, University of Edinburgh, UK
Zornitsa Kozareva, University of Alicante, Spain
Chin-Yew Lin, Microsoft Research Asia, China
Berenike Loos, European Media Laboratory GmbH, Heidelberg, Germany
Bernardo Magnini, ITC-irst, Trento, Italy
Irina Matveeva, University of Chicago, USA
Rada Mihalcea, University of North Texas, USA
Andrea Mulloni, University of Wolverhampton, UK

Roberto Navigli, University of Roma “La Sapienza”, Italy
Malvina Nissim, University of Bologna, Italy
Joakim Nivre, Växjö University and Uppsala University, Sweden
Constantin Orasan, University of Wolverhampton, UK
Rainer Osswald, FernUniversität in Hagen, Germany
Sebastian Padó, Saarland University, Germany
Adam Przepiórkowski, Polish Academy of Sciences, Poland
Reinhard Rapp, Universitat Rovira i Virgili, Tarragona, Spain
Bettina Schrader, University of Osnabrück, Germany
Sabine Schulte im Walde, Universität Stuttgart, Germany
Serge Sharoff, University of Leeds, UK
Yihai Shen, The Hong Kong University of Science and Technology, Hong Kong
Anders Søgaard, University of Copenhagen, Denmark
Lucia Specia, University of São Paulo, Brasil
Joel Tetreault, University of Pittsburgh, USA
Reut Tsarfaty, University of Amsterdam, The Netherlands
Begoña Villada Moirón, University of Groningen, The Netherlands
Stephen Wan, Macquarie University, Australia
Janyce M. Wiebe, University of Pittsburgh, USA
Hans Friedrich Witschel, University of Leipzig, Germany
Bing Zhao, Carnegie Mellon University, USA

Table of Contents

Conference Program

Monday, June 25, 2007

14:45–16:35 Poster Session

Posters

Measuring Syntactic Difference in British English
Nathan C. Sanders

Inducing Combinatory Categorial Grammars with Genetic Algorithms
Elias Ponvert

An Implementation of Combined Partial Parser and Morphosyntactic Disambiguator
Aleksander Buczyński

A Practical Classification of Multiword Expressions
Radosław Moszczyński

Automatic Prediction of Cognate Orthography Using Support Vector Machines
Andrea Mulloni

Exploiting Structure for Event Discovery Using the MDI Algorithm
Martina Naughton

Kinds of Features for Chinese Opinionated Information Retrieval
Taras Zagibalov

Tuesday, June 26, 2007

9:15–9:25 Opening Remarks

Grammar and the Lexicon

09:25–09:50 *Limitations of Current Grammar Induction Algorithms*
Bart Cramer

09:50-10:15 *Logistic Online Learning Methods and Their Application to Incremental Dependency Parsing*
Richard Johansson

10:15-10:40 *Adaptive String Distance Measures for Bilingual Dialect Lexicon Induction*
Yves Scherrer

Quantitative and Formal Linguistics

14:30-14:55 *Identifying Linguistic Structure in a Quantitative Analysis of Dialect Pronunciation*
Jelena Prokić

14:55-15:20 *Towards a Computational Treatment of Superlatives*
Silke Scheible

Semantics

15:45-16:10 *Annotating and Learning Compound Noun Semantics*
Diarmuid Ó Séaghdha

16:10-16:35 *Semantic Classification of Noun Phrases Using Web Counts and Learning Algorithms*
Paul Nulty

16:35-17:00 *Computing Lexical Chains with Graph Clustering*
Olena Medelyan

17:00-17:25 *Clustering Hungarian Verbs on the Basis of Complementation Patterns*
Kata Gábor and Enikő Héja