EACL 2014

14th Conference of the European Chapter of the Association for Computational Linguistics

Proceedings of the Student Research Workshop

April 26-30, 2014
Gothenburg, Sweden

GOLD SPONSORS

SILVER SPONSOR

BRONZE SPONSORS

SUPPORTERS

EXHIBITORS

OTHER SPONSORS

HOSTS

CLT
Centre for Language Technology
Gothenburg, Sweden

©2014 The Association for Computational Linguistics

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

ISBN 978-1-937284-80-0

Introduction

Welcome to the EACL 2014 Student Research Workshop.

This workshop continues the tradition of providing a forum for student researchers and builds on the success of the previous workshops held in Bergen (1999), Toulouse (2001), Budapest (2003), Trento (2006), Athens (2009), and Avignon (2012). It is an excellent venue for student researchers investigating topics in computational linguistics and natural language processing to present and discuss their research, to meet potential advisors and leading world experts in their research fields, as well as to receive feedback from the international research community.

This year we have introduced two different types of submissions: 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 accepted research papers are presented as talks in two separate sessions allocated for the workshop during the main EACL 2014 conference, while all accepted thesis proposals are presented as posters during the main EACL 2014 poster session.

On behalf of the entire Program Committee, we are delighted to present the proceedings of the Student Research Workshop. We received 13 thesis proposals and 29 research papers. We accepted 5 thesis proposals and 8 research papers leading to an acceptance rate of 38% for thesis proposals and 28% for research papers. The overall quality of the submissions was high, and we thank our program committee for their dedicated and thorough work and excellent feedback.

We also thank our faculty advisor Sebastian Padó for his suggestions, feedback and his extremely quick responding to all our questions. We also thank the EACL 2014 organizing committee, especially Shuly Wintner, Stefan Riezler, Sharon Goldwater, Nina Tahmasebi, Gosse Bouma and Yannick Parmentier, for providing us advice and assistance in planning and organizing this workshop. We also want to thank the EACL for providing financial support for students who would otherwise be unable to attend the workshop and the conference.

We truly hope you will enjoy the Student Research Workshop in Gothenburg!

Desmond Elliott, University of Edinburgh
Konstantina Garoufi, University of Potsdam
Douwe Kiela, University of Cambridge
Ivan Vulić, KU Leuven

EACL 2014 Student Research Workshop Co-Chairs

Student Chairs:

Desmond Elliott, *University of Edinburgh*
Konstantina Garoufi, *University of Potsdam*
Douwe Kiela, *University of Cambridge*
Ivan Vulić, *KU Leuven*

Faculty Advisor:

Sebastian Padó, *University of Stuttgart*

Program Committee:

Marianna Apidianaki, *LIMSI-CNRS*
Borja Balle, *McGill University*
Marco Baroni, *University of Trento*
Timo Baumann, *University of Hamburg*
Lee Becker, *Hapara*
Steven Bethard, *University of Alabama at Birmingham*
Chris Biemann, *TU Darmstadt*
Hendrik Buschmeier, *University of Bielefeld*
Marcela Charfuelan, *DFKI*
Christian Chiarcos, *University of Potsdam*
Laurence Danlos, *Université Paris 7*
Jan De Belder, *KU Leuven*
Vladimir Eidelman, *University of Maryland*
Jacob Eistenstein, *Georgia Tech*
Antske Fokkens, *University of Amsterdam*
Goran Glavaš, *University of Zagreb*
João Graça, *Inesc-ID*
Weiwei Guo, *Columbia University*
Dilek Hakkani-Tür, *Microsoft Research*
Bo Han, *University of Melbourne*
Katja Hofmann, *Microsoft Research*
Ann Irvine, *Johns Hopkins University*
David Jurgens, *Sapienza University*
Anna Kazantseva, *University of Ottawa*
Philipp Koehn, *University of Edinburgh*
Oleksandr Kolomiyets, *KU Leuven*
Sebastian Krause, *DFKI*
Vasileios Lampos, *University College London*
Els Lefever, *Ghent University*
Pierre Lison, *University of Oslo*
Elijah Mayfield, *Carnegie Mellon University*
Roser Morante, *University of Antwerp*
Preslav Nakov, *Qatar Computing Research Institute*
Dong Nguyen, *University of Twente*
Joakim Nivre, *Uppsala University*
Gabriella Pasi, *University of Milano Bicocca*

Alexandre Passos, *UMass Amherst*
Andreas Peldszus, *University of Potsdam*
Tamara Polajnar, *University of Cambridge*
Ariadna Quattoni, *UPC*
Sravana Reddy, *Dartmouth College*
Michaela Regneri, *Saarland University*
Christian Scheible, *University of Stuttgart*
Amanda Stent, *Yahoo! Labs*
Kristina Striegnitz, *Union College*
Anders Søgaard, *University of Copenhagen*
Joel Tetreault, *Yahoo! Labs*
Oscar Täckström, *Google*
Tim Van de Cruys, *IRIT*
Svitlana Volkova, *Johns Hopkins University*
Mengqiu Wang, *Stanford University*
Rui Wang, *DFKI*
Jason Williams, *Microsoft Research*
Alessandra Zarcone, *University of Stuttgart*
Jan Šnajder, *University of Zagreb*

Table of Contents

Literature-based discovery for Oceanographic climate science
Elias Aamot 1

Unsupervised Relation Extraction of In-Domain Data from Focused Crawls
Steffen Remus 11

Enhancing Medical Named Entity Recognition with Features Derived from Unsupervised Methods
Maria Skeppstedt 21

Now We Stronger than Ever: African-American English Syntax in Twitter
Ian Stewart 31

Expanding the Range of Automatic Emotion Detection in Microblogging Text
Jasy Suet Yan Liew 38

Resolving Coreferent and Associative Noun Phrases in Scientific Text
Ina Roesiger and Simone Teufel 45

Modelling Irony in Twitter
Francesco Barbieri and Horacio Saggion 56

Multi-class Animacy Classification with Semantic Features
Johannes Bjerva 65

Using Minimal Recursion Semantics for Entailment Recognition
Elisabeth Lien 76

A Graph-Based Approach to String Regeneration
Matic Horvat and William Byrne 85

Complexity of Word Collocation Networks: A Preliminary Structural Analysis
Shibamouli Lahiri 96

Automatic Creation of Arabic Named Entity Annotated Corpus Using Wikipedia
Maha Althobaiti, Udo Kruschwitz and Massimo Poesio 106

Generating artificial errors for grammatical error correction
Mariano Felice and Zheng Yuan 116

Student Research Workshop Program

Monday, April 28, 2014

15:15–18:30 **Posters (Thesis Proposals)**

Literature-Based Discovery for Oceanographic Climate Science
Elias Aamot

Unsupervised Relation Extraction of In-Domain Data from Focused Crawls
Steffen Remus

Enhancing Medical Named Entity Recognition with Features Derived from Unsupervised Methods
Maria Skeppstedt

Now We Stronger than Ever: African-American English Syntax in Twitter
Ian Stewart

Expanding the Range of Automatic Emotion Detection in Microblogging Text
Jasy Suet Yan Liew

Tuesday, April 29, 2014

12:30–14:00 SRW Lunch

16:30–18:10 **Parallel Session I (Research Papers I)**

16:30–16:55 *Resolving Coreferent and Associative Noun Phrases in Scientific Text*
Ina Roesiger and Simone Teufel

16:55–17:20 *Modelling Irony in Twitter*
Francesco Barbieri and Horacio Saggion

17:20–17:45 *Multi-class Animacy Classification with Semantic Features*
Johannes Bjerva

17:45–18:10 *Using Minimal Recursion Semantics for Entailment Recognition*
Elisabeth Lien

16:30–18:10 **Parallel Session II (Research Papers II)**

16:30–16:55 *A Graph-Based Approach to String Regeneration*
Matic Horvat and William Byrne

16:55–17:20 *Complexity of Word Collocation Networks: A Preliminary Structural Analysis*
Shibamouli Lahiri

Tuesday, April 29, 2014 (continued)

16:30–18:10 **Parallel Session II (Research Papers II) (continued)**

17:20–17:45 *Automatic Creation of Arabic Named Entity Annotated Corpus Using Wikipedia*
Maha Althobaiti, Udo Kruschwitz and Massimo Poesio

17:45–18:10 *Generating Artificial Errors for Grammatical Error Correction*
Mariano Felice and Zheng Yuan