RANLPStud 2011

Proceedings of the
Student Research Workshop

associated with

The 8th International Conference on
Recent Advances in Natural Language Processing
(RANLP 2011)

13 September, 2011
Hissar, Bulgaria

STUDENT RESEARCH WORKSHOP
ASSOCIATED WITH THE INTERNATIONAL CONFERENCE
RECENT ADVANCES IN
NATURAL LANGUAGE PROCESSING'2011

PROCEEDINGS

Hissar, Bulgaria
13 September 2011

ISBN 978-954-452-016-8

Designed and Printed by INCOMA Ltd.
Shoumen, BULGARIA

Preface

The Recent Advances in Natural Language Processing (RANLP) conference, already in its eight year and ranked among the most influential NLP conferences, has always been a meeting venue for scientists coming from all over the world. Since 2009, we decided to give arena to the younger and less experienced members of the NLP community to share their results with an international audience. For this reason, further to the first successful and highly competitive Student Research Workshop associated with the conference RANLP 2009, we are pleased to announce the second edition of the workshop which is held during the main RANLP 2011 conference days on 13 September 2011.

The aim of the workshop is to provide an excellent opportunity for students at all levels (Bachelor, Master, and Ph.D.) to present their work in progress or completed projects to an international research audience and receive feedback from senior researchers. We have received 31 high quality submissions, among which 6 papers have been accepted as regular oral papers, and 18 as posters. Each submission has been reviewed by at least 2 reviewers, who are experts in their field, in order to supply detailed and helpful comments. The papers' topics cover a broad selection of resarch areas, such as:

- Annotation;
- BioMedical NLP;
- Coreference Resolution;
- Corpus Linguistics;
- Discourse Processing;
- Information Extraction;
- Machine Translation;
- Ontologies;
- Opinion Mining;
- Natural Language Generation;
- Parsing;
- Part-of-Speech Tagging;
- Question Answering;
- Text Classification;
- Text Segmentation;
- Text Summarization;
- Textual Entailment;
- Word Sense Disambiguation.

We are also glad to admit that our authors comprise a very international group with students coming from: Brazil, Bulgaria, France, Germany, Hungary, India, Iran, Romania, Russia, Spain, Serbia, Sweden, United Kingdom and United States.

We would like to thank the authors for submitting their articles to the Student Workshop and the members of the Programme Committee for their efforts to provide exhaustive reviews and for reacting in time. We are especially grateful to the RANLP Chairs Prof. Galia Angelova and Prof. Ruslan Mitkov for their indispensable support and encouragement during the Workshop organisation.

We hope that all the participants will receive invaluable feedback about their work. This year the conference and the workshop will take place in a new location (Hissar, Bulgaria), so we wish you to enjoy this new location and the Workshop!

Irina Temnikova, Ivelina Nikolova and Natalia Konstantinova
Organisers of the Student Workshop, held in conjunction with
The International Conference RANLP-11

Organizers:

Irina Temnikova (University of Wolverhampton, UK)
Ivelina Nikolova (Bulgarian Academy of Sciences, Bulgaria)
Natalia Konstantinova (University of Wolverhampton, UK)

Programme Committee:

Alexandra Balahur (University of Alicante, Spain)
Chris Biemann (Technical University Darmstadt, Germany)
Kevin Bretonnel Cohen (University of Colorado School of Medicine, USA)
Iustin Dornescu (University of Wolverhampton, UK)
Atefeh Farzindar (NLP Technologies Inc., Canada)
Darja Fišer (University of Ljubljana, Slovenia)
Najeh Hajlaoui (University of Wolverhampton, UK)
Laura Hasler (University of Strathclyde, UK)
Iustina Ilisei (University of Wolverhampton, UK)
Diana Inkpen (University of Ottawa, Canada)
Sobha Lalitha Devi (AU-KBC Research Centre, India)
Nikola Ljubešić (University of Zagreb, Croatia)
Wolfgang Maier (University of Düsseldorf, Germany)
Preslav Nakov (National University of Singapore, Singapore)
Constantin Orasan (University of Wolverhampton, UK)
Petya Osenova (Bulgarian Academy of Sciences, Bulgaria)
George Paltoglou (University of Wolverhampton, UK)
Ivandre Paraboni (University of Sao Paulo, Brazil)
Marta Recasens (University of Barcelona, Spain)
Georg Rehm (DFKI, Berlin, Germany)
Miguel Rios Gaona (University of Wolverhampton, UK)
Raphaël Rubino (University of Avignon, France)
Sebastian Rudolph (Karlsruher Institut für Technologie, Germany)
Doaa Samy (University of Cairo, Egypt)
Luis Sarmento (University of Porto, Portugal)
Thamar Solorio (University of Alabama at Birmingham, USA)
Lucia Specia (University of Wolverhampton, UK)
Asher Stern (Bar-Ilan University, Israel)
Ang Sun (New York University, USA)
Cristina Toledo (University of Málaga, Spain)
Yoshimasa Tsuruoka (Japan Advanced Institute of Science and Technology, Japan)
Cristina Vertan (University of Hamburg, Germany)
Pinar Wennerberg (Bayer, Germany)
Wajdi Zaghouani (University of Pennsylvania, USA)
Torsten Zesch (Technical University Darmstadt, Germany)

Table of Contents

Workshop Programme

Tuesday, 13 September, 2011

10:00–10:05 Opening

PLOVDIV hall: Oral Presentations

10:05–10:25 *Domain-Dependent Detection of Light Verb Constructions*
István T. Nagy, Gábor Berend, György Móra and Veronika Vincze

10:25–10:45 *A Weighted Lexicon of French Event Names*
Béatrice Arnulphy

11:00–11:30 Coffee Break and Student Posters (Lobby)

HISSAR hall: Oral Presentations

11:30–11:50 *Towards a Better Exploitation of the Brown 'Family' Corpora in Diachronic Studies of British and American English Language Varieties*
Sanja Štajner

11:50–12:10 *Projecting Farsi POS Data To Tag Pashto*
Mohammad Khan, Eric Baucom, Anthony Meyer and Lwin Moe

12:10–12:30 *Enriching Phrase-Based Statistical Machine Translation with POS Information*
Miriam Kaeshammer and Dominikus Wetzel

12:30–12:50 *Inter-domain Opinion Phrase Extraction Based on Feature Augmentation*
Gábor Berend, István T. Nagy, György Móra and Veronika Vincze

Lobby: Poster Presentations

15:40–16:20

ArbTE: Arabic Textual Entailment
Maytham Alabbas

RDFa Editor for Ontological Annotation
Melania Duma

Extracting Protein-Protein Interactions with Language Modelling
Ali Reza Ebadat

Experiments with Small-size Corpora in CBMT
Monica Gavrila and Natalia Elita

Question Parsing for QA in Spanish
Iria Gayo

Tuesday, 13 September, 2011 (continued)