RANLPStud 2013

Proceedings of the Student Research Workshop

associated with

The 9th International Conference on Recent Advances in Natural Language Processing (RANLP 2013)

9–11 September, 2013
Hissar, Bulgaria

STUDENT RESEARCH WORKSHOP
ASSOCIATED WITH THE INTERNATIONAL CONFERENCE
RECENT ADVANCES IN
NATURAL LANGUAGE PROCESSING'2013

PROCEEDINGS

Hissar, Bulgaria
9–11 September 2013

ISSN 1314-9156

Designed and Printed by INCOMA Ltd.
Shoumen, BULGARIA

Preface

The Recent Advances in Natural Language Processing (RANLP) conference, which is 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 and second successful and highly competitive Student Research Workshops associated with the conference RANLP 2009 and RANLP 2011, we are pleased to announce the third edition of the workshop which is held during the main RANLP 2013 conference days, 9–11 September 2013.

The aim of the workshop is to provide an excellent opportunity for students at all levels (Bachelor, Masters, and Ph.D.) to present their work in progress or completed projects to an international research audience and receive feedback from senior researchers. We received 36 high quality submissions, among which 4 papers have been accepted for oral presentation, 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 research areas, such as:

- application-orientated papers related to NLP;
- computer-aided language learning;
- dialogue systems;
- discourse;
- electronic dictionaries;
- evaluation;
- information extraction, event extraction, term extraction;
- information retrieval;
- knowledge acquisition;
- language resources, corpora, terminologies;
- lexicon;
- machine translation;
- morphology, syntax, parsing, POS tagging;
- multilingual NLP;
- NLP for biomedical texts;
- NLP for the Semantic web;
- ontologies;
- opinion mining;
- question answering;
- semantic role labelling;
- semantics;
- speech recognition;
- temporality processing;
- text categorisation;
- text generation;
- text simplification and readability estimation;
- text summarisation;
- textual entailment;
- theoretical papers related to NLP;
- word-sense disambiguation;

We are also glad to admit that our authors comprise a very international group with students coming from: Belgium, China, Croatia, France, Germany, India, Italy, Luxembourg, Russian Federation, Spain, Sweden, Tunisia and the United Kingdom.

We would like to thank the authors for submitting their articles to the Student Workshop, the members of the Programme Committee for their efforts to provide exhaustive reviews, and the mentors who agreed to have a deeper look at the students' work. We hope that all the participants will receive invaluable feedback about their research.

Irina Temnikova, Ivelina Nikolova and Natalia Konstantinova
Organisers of the Student Workshop, held in conjunction with
The International Conference RANLP-13

Organizers:

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

Program Committee:

Chris Biemann (Technical University Darmstadt, Germany)
Kevin Bretonnel Cohen (University of Colorado School of Medicine, USA)
Iustin Dornescu (University of Wolverhampton, UK)
Laura Hasler (University of Wolverhapton, UK)
Diana Inkpen (University of Ottawa, Canada)
Natalia Konstantinova (University of Wolverhampton, UK)
Sebastian Krause (DFKI, Germany)
Sandra Kübler (University of Indiana, USA)
Lori Lamel (CNRS/LIMSI, France)
Annie Louis (University of Edinburgh, UK)
Preslav Nakov (QCRI, Qatar)
Ivelina Nikolova (Bulgarian Academy of Sciences, Bulgaria)
Constantin Orasan (University of Wolverhampton, UK)
Petya Osenova (Sofia University and IICT-BAS, Bulgaria)
Ivandre Paraboni (University of Sao Paulo, Brazil)
Michael Poprat (Averbis GmbH, Freiburg)
Rashmi Prasad (University of Wisconsin-Milwaukee, USA)
Raphael Rubino (Dublin City University and Symantec, Ireland)
Doaa Samy (Autonoma University of Madrid, Spain)
Thamar Solorio (University of Alabama at Birmingham, USA)
Stan Szpakowicz (University of Ottawa, Canada)
Irina Temnikova (Bulgarian Academy of Sciences, Bulgaria)
Eva Maria Vecchi (University of Trento, Italy)
Cristina Vertan (University of Hamburg, Germany)
Feiyu Xu (DFKI, Germany)
Torsten Zesch (Technical University Darmstadt, Germany)

Table of Contents

Workshop Programme

Monday September 9, 2013 (16:30 - 18:30)

Methods, Resources and Language Processing Tasks (Posters Session 1)

A Dataset for Arabic Textual Entailment
Maytham Alabbas

Automatic Evaluation of Summary Using Textual Entailment
Pinaki Bhaskar and Partha Pakray

Detecting Negated and Uncertain Information in Biomedical and Review Texts
Noa Cruz

Cross-Language Plagiarism Detection Methods
Vera Danilova

Random Projection and Geometrization of String Distance Metrics
Daniel Hromada

Improving Language Model Adaptation using Automatic Data Selection and Neural Network
Shahab Jalalvand

Towards Basque Oral Poetry Analysis: A Machine Learning Approach
Mikel Osinalde, Aitzol Astigarraga, Igor Rodriguez and Manex Agirrezabal

Reporting Preliminary Automatic Comparable Corpora Compilation Results
Ekaterina Stambolieva

Tuesday September 10, 2013 (11:30 - 12:50)

Oral Presentations

Event-Centered Simplification of News Stories
Goran Glavaš and Sanja Štajner

Named Entity Recognition in Broadcast News Using Similar Written Texts
Niraj Shrestha and Ivan Vulić

Collection, Annotation and Analysis of Gold Standard Corpora for Knowledge-Rich Context Extraction in Russian and German
Anne-Kathrin Schumann

Unsupervised Learning of A-Morphous Inflection with Graph Clustering
Maciej Janicki

Wednesday September 11, 2013 (15:30 - 17:30)

Applications (Poster Session 2)

Perceptual Feedback in Computer Assisted Pronunciation Training: A Survey
Renlong Ai

Answering Questions from Multiple Documents –
the Role of Multi-Document Summarization
Pinaki Bhaskar

Multi-Document Summarization using Automatic Key-Phrase Extraction
Pinaki Bhaskar

Towards a Discourse Model for Knowledge Elicitation
Eugeniu Costetchi

Rule-based Named Entity Extraction for Ontology Population
Aurore De Amaral

Towards Definition Extraction Using Conditional Random Fields
Luis Espinosa Anke

Statistical-based System for Morphological Annotation of Arabic Texts
Nabil Khoufi and Manel Boudokhane

A System for Generating Cloze Test Items from Russian-Language Text
Andrey Kurtasov

GF Modern Greek Resource Grammar
Ioanna Papadopoulou

Korean Word-Sense Disambiguation Using Parallel Corpus as Additional Resource
Chungen Li