RANLPStud 2015

Proceedings of the Student Research Workshop

associated with

The 10th International Conference on Recent Advances in Natural Language Processing (RANLP 2015)

7–9 September, 2015
Hissar, Bulgaria

STUDENT RESEARCH WORKSHOP
ASSOCIATED WITH THE INTERNATIONAL CONFERENCE
RECENT ADVANCES IN
NATURAL LANGUAGE PROCESSING'2015

PROCEEDINGS

Hissar, Bulgaria
7–9 September 2015

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 three successful and highly competitive Student Research Workshops associated with the conferences RANLP 2009, RANLP 2011, and RANLP 2013, we are pleased to announce the forth edition of the workshop which is held during the main RANLP 2015 conference days, 7–9 September 2015.

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. This year, we received 10 high quality submissions, among which 3 papers have been accepted for oral presentation, and 2 as posters. Each submission has been reviewed by at least 3 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;

As usual, our authors comprise a large international group with students coming from: Bulgaria, Germany, India, 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 Alexander Popov
Organisers of the Student Workshop, held in conjunction with
The International Conference RANLP-15

Organizers:

Irina Temnikova (Qatar Computing Research Institute, HBKU, Qatar)
Ivelina Nikolova (Bulgarian Academy of Sciences, Bulgaria)
Alexander Popov (Bulgarian Academy of Sciences, Bulgaria)

Programme Committee:

Ahmed Abdelali (Qatar Computing Research Institute, HBKU, Qatar)
Corina Forascu (Alexandru Ioan Cuza University, Romania)
Goran Glavaš (University of Zagreb, Croatia)
Francisco Guzmán (Qatar Computing Research Institute, HBKU, Qatar)
Diana Inkpen (University of Ottawa, Canada)
Annie Louis (University of Edinburgh, United Kingdom)
Preslav Nakov (Qatar Computing Research Institute, HBKU, Qatar)
Constantin Orasan (University of Wolverhampton, UK)
Petya Osenova (Sofia University and IICT-BAS, Bulgaria)
Doaa Samy (Cairo University, Egypt)
Jan Šnajder (University of Zagreb, Croatia)
Stan Szpakowicz (University of Ottawa, Canada)
Andrea Varga (The Content Group Ltd, United Kingdom)

Table of Contents

The Complexity of Scrambling in Japanese: A TAG Approach
Alexander Diez 1

Some Theoretical Considerations in Off-the-Shelf Text Analysis Software
Emma Franklin 8

Unsupervised Topic-Specific Domain Dependency Graphs for Aspect Identification in Sentiment Analysis
Sarah Kohail 16

Unsupervised learning of agglutinated morphology using nested Pitman-Yor process based morpheme induction algorithm
Arun Kumar 24

Easy-read Documents as a Gold Standard for Evaluation of Text Simplification Output
Victoria Yaneva 30