RANLPStud 2019

Proceedings of the Student Research Workshop

associated with

The 12th International Conference on Recent Advances in Natural Language Processing (RANLP 2019)

2–4 September, 2019
Varna, Bulgaria

STUDENT RESEARCH WORKSHOP
ASSOCIATED WITH THE INTERNATIONAL CONFERENCE
RECENT ADVANCES IN
NATURAL LANGUAGE PROCESSING'2019

PROCEEDINGS

Varna, Bulgaria
2–4 September 2019

Series Online ISSN 2603-2821

Designed and Printed by INCOMA Ltd.
Shoumen, BULGARIA

Preface

The RANLP 2019 Student Research Workshop (RANLPStud 2019) is a special track of the established international conference Recent Advances in Natural Language Processing (RANLP 2019), now in its twelfth edition.

The RANLP Student Research Workshop is being organised for the sixth time and this year is running in parallel with the other tracks of the main RANLP 2019 conference. The target of RANLPStud 2019 is to be a discussion forum and provide an outstanding 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.

The RANLP 2019 Student Research Workshop received a large number of submissions (23), a fact which was reflecting the record number of events, sponsors, submissions, and participants at the main RANLP 2019 conference.

We have accepted 2 excellent student papers as oral presentations and 12 submissions will be presented as posters. The final acceptance rate of the workshop was 60%.

We made our best to make the reviewing process in the best interest of our authors, by asking our reviewers to give as most exhaustive comments and suggestions as possible as well as to maintain an encouraging attitude. Each student submission was reviewed by 2 to 3 Programme Committee members, which are specialists in their field and were carefully selected to match the submission's topic.

This year, as usual, we invited both strictly Natural Language Processing (NLP) submissions, and submissions at the borderline between two sciences (but bearing contributions to NLP. The topics of the accepted submissions include:

- Adding Linguistic Knowledge to NLP Tasks
- Multilingual Complex Word Identification
- Text Normalisation
- Text Classification for Social Media
- Automatic Speech Recognition
- Named Entity Recognition
- Part-Of-Speech Tagging
- Cross-Lingual Coreference
- Content-Based Recommender Systems for Books
- Corpora and Processing Tools
- Question Answering, Machine Reading Comprehension

Our authors have a rich variety of nationalities and affiliation countries: Bulgaria, USA, Russia, India, Kazakhstan, Germany, Ireland, Switzerland, Saudi Arabia and Spain.

We are thankful to the members of the Programme Committee for having provided such exhaustive reviews and even accepting additional reviews, and to the conference mentors, who provided additional comments to participants.

Venelin Kovatchev, Irina Temnikova, Branislava Šandrih, and Ivelina Nikolova

Organisers of the Student Research Workshop, special track of the International Conference RANLP 2019

Organizers:

Venelin Kovatchev (Universitat de Barcelona)
Irina Temnikova (freelancer)
Branislava Šandrih (Belgrade University)
Ivelina Nikolova (Bulgarian Academy of Sciences, Bulgaria, Ontotext AD)

Programme Committee:

Aleksandar Savkov (Babylon Health)
Corina Forascu (University "Al. I. Cuza" Iaşi)
Darina Gold (University of Duisburg-Essen)
Diego Moussallem (Paderborn University)
Dmitry Ilvovsky (Higher School of Economics)
Eduardo Blanco (University of North Texas)
Horacio Rodriguez (Universitat Politècnica de Catalunya)
Josef Steinberger (University of West Bohemia)
Liviu P. Dinu (University of Bucharest)
M. Antonia Marti (Universitat de Barcelona)
Mariona Taulé (Universitat de Barcelona)
Navid Rekabsaz (Idiap Research Institute)
Paolo Rosso (Universitat Politècnica de Valencia)
Sandra Kübler (Indiana University Bloomington)
Shervin Malmasi (Harvard Medical School)
Sobha Lalitha Devi (AU-KBC Research Centre)
Thierry Declerck (DFKI GmbH)
Tracy Holloway King (A9.com, Stanford University)
Vijay Sundar Ram (AU-KBC Research Centre, MIT Campus of Anna University)
Yannis Korkontzelos (Edge Hill University)

Table of Contents

Normalization of Kazakh Texts
Assina Abdussaitova and Alina Amangeldiyeva 1

Classification Approaches to Identify Informative Tweets
Piush Aggarwal 7

Dialect-Specific Models for Automatic Speech Recognition of African American Vernacular English
Rachel Dorn 16

Multilingual Language Models for Named Entity Recognition in German and English
Antonia Baumann 21

Parts of Speech Tagging for Kannada
Swaroop L R, Rakshith Gowda G S, Sourabh U and Shriram Hegde 28

Cross-Lingual Coreference: The Case of Bulgarian and English
Zara Kancheva 32

Towards Accurate Text Verbalization for ASR Based on Audio Alignment
Diana Geneva and Georgi Shopov 39

Evaluation of Stacked Embeddings for Bulgarian on the Downstream Tasks POS and NERC
Iva Marinova 48

Overview on NLP Techniques for Content-based Recommender Systems for Books
Melania Berbatova 55

Corpora and Processing Tools for Non-standard Contemporary and Diachronic Balkan Slavic
Teodora Vukovic, Nora Muheim, Olivier Winistörfer, Ivan Šimko, Anastasia Makarova and Sanja Bradjan 62

Question Answering Systems Approaches and Challenges
Reem Alqifari 69

Adding Linguistic Knowledge to NLP Tasks for Bulgarian: The Verb Paradigm Patterns
Ivaylo Radev 76

Multilingual Complex Word Identification: Convolutional Neural Networks with Morphological and Linguistic Features
Kim Cheng SHEANG 83

Neural Network-based Models with Commonsense Knowledge for Machine Reading Comprehension
Denis Smirnov 90