INTERNATIONAL CONFERENCE
RECENT ADVANCES
IN NATURAL LANGUAGE PROCESSING

R A N L P 2 0 2 3

Large Language Models for Natural Language Processing

P R O C E E D I N G S

Edited by Galia Angelova, Maria Kunilovskaya and Ruslan Mitkov

Varna, Bulgaria
4–6 September, 2023
https://ranlp.org/ranlp2023

INTERNATIONAL CONFERENCE
RECENT ADVANCES IN
NATURAL LANGUAGE PROCESSING 2023

Large Language Models
for Natural Language Processing

PROCEEDINGS

4–6 September, 2023
https://ranlp.org/ranlp2023

Online ISBN 978-954-452-092-2
e-book site: www.acl-bg.org
Series Online ISSN 2603-2813.2023

INCOMA Ltd.
Shoumen, BULGARIA

Preface

The international RANLP conference is a well-established biennial forum for computational linguists and Natural Language Processing (NLP) practitioners which continues to report important trends in the field. In 2023 the programme was dominated by research on developing or exploiting pre-trained large language models (LLMs) and the deep learning technology which was the reason to assign the subtitle 'LLMs for NLP' to the volume. The highlights of this year included the urgent and challenging topics such as responsible and explainable machine learning, quality of the existing datasets, multimodality and multilinguality.

The conferences attracted 165 submissions and accepted 31 regular papers, 59 short papers, 41 posters, and 4 demos (excluding workshops). The event was attended by over 170 participants from over 35 countries.

The conference in 2023 features six keynote speakers:

- Eduard Hovy (University of Melbourne, Australia and Carnegie Mellon University, USA),
- Tharindu Ranasinghe (Aston University, UK),
- Sandra Kübler (Indiana University Bloomington, USA),
- Lucas Beyer (Google Brain, Switzerland),
- Isabelle Augenstein (University of Copenhagen, Denmark),
- Efstathios Stamatatos (University of the Aegean, Greece).

The proceedings cover a wide variety of NLP topics, including training, adaptation, evaluation and explanation of language models, multimodal studies, language resources, machine translation, NLP for social sciences and literary studies, simplification and summarisation, topic modelling, opinion-mining and sentiment analysis, fake news, bias and hate speech detection.

In 2023 RANLP was preceded by the summer school 'Deep Learning for NLP' and pre-conference tutorials, and hosted a record number of post-conference workshops on popular NLP topics:

- LT-EDI 2023 – Third Workshop on Language Technology for Equality, Diversity and Inclusion
- DravidianLangTech 2023 – Third Workshop on Speech and Language Technologies for Dravidian languages
- TSAR 2023 – Workshop on Text Simplification, Accessibility and Readability
- ALP 2023 – Workshop on Ancient Language Processing
- HumEval 2023 – Third Workshop on Human Evaluation of NLP Systems
- BUCC 2023 – 16th Workshop on Building and Using Comparable Corpora
- CASE 2023 – 6th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text
- ConTeNTS 2023 – Computational Terminology in NLP and Translation Studies
- NLP4TIA 2023 – NLP tools and resources for translation and interpreting applications

In addition to thanking the keynote speakers and workshop organisers who accepted our invitation, we would like to thank the lecturers and tutors of the Summers school and tutorials.

We are grateful to the members of the Programme Committee and all additional reviewers. They ensured that the best papers were included in the Proceedings and provided invaluable comments to the authors.

We would like to use this paragraph to acknowledge the members of the Organising Committee, who worked very hard during the last few months and whose dedication and efforts made the organisation of this event possible. The members of the Organising Committee (listed in alphabetical order below) carried out numerous organisational tasks and were eager to step in and support the organisation of the conference whenever needed: Khadija Ait ElFqih, Elena Blagoeva, Marie Escribe, Emma Franklin, Amal Haddad Haddad, Jessica López Espejel, Teodora Mihajlov and Nikolai Nikolov.

A big THANK YOU to all of you, this conference could not have taken place so smoothly without you!

Finally, many thanks go to Lancaster University and the Institute of Information and Communication Technologies at the Bulgarian Academy of Sciences for their unreserved support of RANLP. Our gratitude goes also to our generous sponsors as well: Bulgarian National Research Fund, Ontotext, Iris.AI, Senso, Cambridge University Press and ELDA.

Varna, 9 September 2023
Galia Angelova, Maria Kunilovskaya and Ruslan Mitkov

The International Conference RANLP-2023 is organised by:

Lancaster University, UK

Institute of Information and Communication Technologies (IICT),
Bulgarian Academy of Sciences, Bulgaria

Sponsors:

Grant КП-06-МНФ/3,
19.05.2023

Programme Committee Chair:
Ruslan Mitkov, Lancaster University, UK

Organising Committee Chair:
Galia Angelova, IICT, Bulgarian Academy of Sciences, Bulgaria

Publishing Team:
Maria Kunilovskaya and Nikolai Nikolov

Programme Committee and Proceedings Coordinators:
Khadija Ait ElFqih, University of Naples l'Orientale, Italy
Elena Blagoeva, Bulgarian Academy of Sciences, Bulgaria
Marie Escribe, Polytechnic University of Valencia & LanguageWire, Spain
Emma Franklin, Renato Software Ltd., United Kingdom
Amal Haddad Haddad, Universidad de Granada, Spain
Maria Kunilovskaya, University of Saarland, Germany
Jessica López Espejel, Novelis, France
Teodora Mihajlov, Univeristy of Belgrade, Serbia
Nikolai Nikolov, INCOMA Ltd., Shoumen, Bulgaria

Programme Committee:

Cengiz Acarturk, Cognitive Science Department, Jagiellonian University, Poland
Heike Adel, Bosch Center for Artificial Intelligence, Germany
Itziar Aldabe, HiTZ Center – Ixa, University of the Basque Country (UPV/EHU), Spain
Hassina Aliane, Research Center On Scientific and Technical Information, Algeria
Galia Angelova, Institute of Information and communication Technologies, Bulgarian Academy of Sciences, Bulgaria
Eduard Barbu, Institute of Computer Science, Estonia
Verginica Barbu, Mititelu RACAI, Romania
Alberto Barrón-Cedeño, Università di Bologna, Italy
Svetla Boytcheva, Ontotext, Bulgaria
António Branco, University of Lisbon, Portugal
Chris Brew, Lexis Nexis, United States
Aljoscha Burchardt, DFKI, Germany
Burcu Can, University of Stirling, United Kingdom
Maya Carrillo, BUAP, Mexico
Kenneth Church, Northeastern University, United States
Jean-Pierre Colson, University of Louvain, Belgium
Mahmoud El-Haj, Lancaster University, United Kingdom
Antonio Ferrandez Rodriguez, University of Alicante, Spain
Emma Franklin, Renato Software Ltd., United Kingdom
Aina Garí Soler, LTCI, Télécom-Paris, Institut Polytechnique de Paris, France
Elizaveta Goncharova, NRU HSE, Russia
Diana Inkpen, University of Ottawa, Canada
Dimitar Kazakov, Department of Computer Science, University of York, United Kingdom
Tracy Holloway, King Adobe Inc., United States
Udo Kruschwitz, University of Regensburg, Germany
Sandra Kübler, Indiana University ,United States
Maria Kunilovskaya, Saarland University, Germany
Els Lefever, LT3, Ghent University, Belgium
Jessica Lopez Espejel, Novelis, France
Natalia Loukachevitch, Lomonosov Moscow State University, Russia
Eugenio Martínez Cámara, University of Jaén, Spain
Diana Maynard, University of Sheffield, United Kingdom
Andres Montoyo, University of Alicante, Spain
Rafael Muñoz Guillena, University of Alicante, Spain
Kemal Oflazer, Carnegie Mellon University, United States
Maciej Ogrodniczuk, Institute of Computer Science, Polish Academy of Sciences, Poland
Antoni Oliver, Universitat Oberta de Catalunya, Spain
Constantin Orăsan, University of Surrey, United Kingdom
John E. Ortega, Northeastern University, United States
Petya Osenova, Sofia University "St. Kl. Ohridski" and IICT-BAS, Bulgaria
Sebastian Padó, Stuttgart University, Germany
Pavel Pecina, Charles University, Czech Republic
David Pinto, Benemérita Universidad Autónoma de Puebla, Mexico

Jakub Piskorski, Polish Academy of Sciences, Poland
Flor Miriam Plaza-del-Arco, Bocconi University, Italy
Alistair Plum, University of Luxembourg, Luxembourg
Massimo Poesio, Queen Mary University of London, United Kingdom
Maja Popović, ADAPT, Dublin City University, Ireland
Jelena Prokic, Leiden University, Netherlands
Gabor Proszeky, Hungarian Research Centre for Linguistics, Hungary
Tharindu Ranasinghe, Aston University, United Kingdom
Paul Rayson, Lancaster University, United Kingdom
Ayla Rigouts Terryn, KU Leuven KULAK, Belgium
Omid Rohanian, University of Oxford, United Kingdom
Branislava Šandrih Todorović, NLB DigIT, Serbia
Raheem Sarwar, OTEHM, Manchester Metropolitan University, United Kingdom
Frederique Segond, Inria & Inalco, France
Nasredine Semmar, CEA LIST, France
Grigori Sidorov, CIC-IPN, Mexico
Gerardo Sierra Martínez, Universidad Nacional Autónoma de México, Mexico
Khalil Simaan, ILLC, University of Amsterdam, Netherlands
Rui Sousa-Silva, University of Porto – Faculty of Arts, Portugal
Felix Stahlberg, Google Research, Germany
Sanja Stajner, Independent Researcher, Germany
Stan Szpakowicz, EECS, University of Ottawa, Canada
Elena Isabelle Tamba, Romanian Academy, Romania
Hristo Tanev, Joint Research Centre, European Commission, Italy
Shiva Taslimipoor, University of Cambridge, United Kingdom
L. Alfonso Ureña-López, University of Jaen, Spain
Victoria Yaneva, National Board of Medical Examiners, United States
Roman Yangarber, University of Helsinki, Finland
Marcos Zampieri, George Mason University ,United States
Michael Zock, CNRS-LIS, France

Additional Reviewers:

Tosin Adewumi, Luleå University of Technology, Sweden
Jacopo Amidei, The Open University, United Kingdom
Isuri Anuradha, University of Wolverhampton, United Kingdom
Santiago Arróniz, Indiana University, United States
Ekaterina Artemova, LMU Munich, Germany
Charmaine Barker, University of York, United Kingdom
Leonor Becerra, Aix-Marseille University, France
Necva Bölücü, Department of Computer Engineering, Hacettepe University, Turkey
Antonina Bondarenko, Université Paris Cité, France
Alba Bonet Jover, University of Alicante, Spain
Pablo Calleja, Universidad Politécnica de Madrid, Spain
Kai Cao, Morgan Stanley, United States
Rémi Cardon, CENTAL, ILC, Université Catholique de Louvain, Belgium
Maria Carmela Cariello, University of Wolverhampton, United Kingdom
Thiago Castro Ferreira, Federal University of Minas Gerais, Brazil
Yue Chen, Microsoft, United States

Daniel Dakota, Indiana University, United States
Angelo Mario Del Grosso, Istituto di Linguistica Computazionale, Consiglio Nazionale delle Ricerche, Italy
Maria Pia di Buono, University of Naples "L'Orientale", Italy
Anna Beatriz Dimas Furtado, University of Galway, Ireland
Marie Escribe, Polytechnic University of Valencia & LanguageWire, Spain
Isabel Espinosa, Zaragoza University of Alicante, Spain
Adam Funk, University of Sheffield, United Kingdom
Dario Garigliotti, University of Bergen, Norway
Federico Gaspari, Dipartimento di Scienze Politiche, Universita' degli Studi di Napoli Federico II, Italy
Carlos Golvano, UniversidadPolitécnicade Madrid, Spain
Le An Ha, RGCL, RIILP, University of Wolverhampton, United Kingdom
Momchil Hardalov, AWS AI Labs, Spain
Nils Hjortnaes, Indiana University Bloomington, United States
Dean Hunter, University of Wolverhampton, United Kingdom
Adrian Iftene, Alexandru Ioan Cuza University of Iasi, Faculty of Computer Science, Romania
Jose Ignacio Abreu Salas, Universidad de Alicante, Spain
Dmitry Ilvovsky, National Research University Higher School of Economics, Russia
Tomoya Iwakura, Fujitsu, Japan
Arkadiusz Janz, Wroclaw University of Science and Technology, Poland
Olha Kanishcheva, University of Jena Germany
Alfiya Khabibullina, University of Malaga, Spain
Nouran Khallaf, University of Leeds, United Kingdom
Lilit Kharatyan, University of Würzburg, Germany
Saranya Krishnamoorthy, Evernorth Health Services, United States
Sobha Lalitha Devi, AU-KBC Research Centre, Anna University, India
Elpida Loupaki, Aristotle University of Thessaloniki, Greece
Giacomo Magnifico, University of Tartu, Estonia
Aaron Maladry, Ghent University, Belgium
Stefano Marchesin, University of Padua, Italy
Michał Marcińczuk, Wrocław University of Science and Technology, Poland
Patricia Martín-Chozas, Ontology Engineering Group, Universidad Politécnica de Madrid
Mikaela Martins, UNISINOS University, Brazil
Irina Matveeva, Reveal, United States
Ivan Vladimir Meza Ruiz, Insituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Mexico
Teodora Mihajlov, University of Belgrade, Serbia
Manuel Montes, INAOE, Mexico
Paloma Moreda Pozo, University of Alicante, Spain
Yasmin Moslem, ADAPT Centre, Dublin City University, Ireland
Moriah Obaje, PhD Student, United Kingdom
Reynier Ortega Bueno, PRHLT Research Center, UPV, Spain
Ondřej Pražák, University of West Bohemia in Pilsen, Czech Republic
Damith Premasiri, University of Wolverhampton, United Kingdom
Prokopis Prokopidis, ILSP/Athena RC, Greece

Marko Putnikovic, University of Belgrade, Serbia
Haizhou Qu, Shenzhen, Polytechnic University, China
Francesco Saina, SSML Carlo Bo, Italy
Robiert Sepulveda Torres, University of Alicante, Spain
Matthew Shardlow, Manchester Metropolitan University, United Kingdom
Prasham Sheth, SLB Software Technology Innovation Center, United States
Archchana Sindhujan, University of Surrey, United Kingdom
Giulia Speranza, University of Naples "L'Orientale", Italy
Arvind Krishna Sridhar, Qualcomm Technologies R&D, United States
Kenneth Steimel, Educational Testing Service, United States
Tianda Sun, University of York, United Kingdom
Colin Swaelens, Ghent University, Belgium
Luigi Talamo, Saarland University, Germany
Antonio Tamayo, Instituto Politécnico Nacional, CIC, Mexico
Zuoyu Tian, Indiana University, United States
Elena Tutubalina, Kazan Federal University, Russia
Rodrigo Wilkens, Université catholique de Louvain, Belgium
Alisa Zhila, Amazon, United States
He Zhou, Indiana University, United States
Inès Zribi, ANLP Research group, MIRACL Lab., Monastir University, Tunisia

Table of Contents