

**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 Kuniorskaya 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, LTCl, 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 Ofazer, 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, Universidad Politécnica de 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, Instituto 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

<i>Bipol: Multi-Axes Evaluation of Bias with Explainability in Benchmark Datasets</i> Tosin Adewumi, Isabella Södergren, Lama Alkhaled, sana al-azzawi, Foteini Simistira Liwicki and Marcus Liwicki	1
<i>Automatically Generating Hindi Wikipedia Pages Using Wikidata as a Knowledge Graph: A Domain-Specific Template Sentences Approach</i> Aditya Agarwal and Radhika Mamidi	11
<i>Cross-lingual Classification of Crisis-related Tweets Using Machine Translation</i> Shareefa Al Amer, Mark Lee and Phillip Smith	22
<i>Lexicon-Driven Automatic Sentence Generation for the Skills Section in a Job Posting</i> Vera Aleksic, Mona Brems, Anna Mathes and Theresa Bertele	32
<i>Multilingual Racial Hate Speech Detection Using Transfer Learning</i> Abinew Ali Ayele, Skadi Dinter, Seid Muhie Yimam and Chris Biemann	41
<i>Exploring Amharic Hate Speech Data Collection and Classification Approaches</i> Abinew Ali Ayele, Seid Muhie Yimam, Tadesse Destaw Belay, Tesfa Asfaw and Chris Biemann	49
<i>Bhojpuri WordNet: Problems in Translating Hindi Synsets into Bhojpuri</i> Imran Ali and Praveen Gatla	60
<i>3D-EX: A Unified Dataset of Definitions and Dictionary Examples</i> Fatemah Almeman, Hadi Sheikhi and Luis Espinosa Anke	69
<i>Are You Not moved? Incorporating Sensorimotor Knowledge to Improve Metaphor Detection</i> Ghadi Alnafesah, Phillip Smith and Mark Lee	80
<i>HAQA and QUQA: Constructing Two Arabic Question-Answering Corpora for the Quran and Hadith</i> Sarah Alnefaie, Eric Atwell and Mohammad Ammar Alsalka	90
<i>ConflBERT-Arabic: A Pre-trained Arabic Language Model for Politics, Conflicts and Violence</i> Sultan Alsarra, Luay Abdeljaber, Wooseong Yang, Niamat Zawad, Latifur Khan, Patrick Brandt, Javier Osorio and Vito D’Orazio	98
<i>A Review in Knowledge Extraction from Knowledge Bases</i> Fabio Yanez, Andrés Montoyo, Yoan Gutierrez, Rafael Muñoz and Armando Suarez	109
<i>Evaluating of Large Language Models in Relationship Extraction from Unstructured Data: Empirical Study from Holocaust Testimonies</i> Isuri Anuradha, Le An Ha, Ruslan Mitkov and Vinita Nahar	117
<i>Impact of Emojis on Automatic Analysis of Individual Emotion Categories</i> Ratchakrit Arreerard and Scott Piao	124
<i>Was That a Question? Automatic Classification of Discourse Meaning in Spanish</i> Santiago Arróniz and Sandra Kübler	132
<i>Designing the LECOR Learner Corpus for Romanian</i> Ana Maria Barbu, Elena Irimia, Carmen Mîrzea Vasile and Vasile Păiș	143

<i>Non-Parametric Memory Guidance for Multi-Document Summarization</i>	
Florian Baud and Alex Aussem	153
<i>Beyond Information: Is ChatGPT Empathetic Enough?</i>	
Ahmed Belkhir and Fatiha Sadat	159
<i>Using Wikidata for Enhancing Compositionality in Pretrained Language Models</i>	
Meriem Beloucif, Mihir Bansal and Chris Biemann	170
<i>Multimodal Learning for Accurate Visual Question Answering: An Attention-Based Approach</i>	
Jishnu Bhardwaj, Anurag Balakrishnan, Satyam Pathak, Ishan Unnarkar, Aniruddha Gawande and Benyamin Ahmadnia	179
<i>Generative Models For Indic Languages: Evaluating Content Generation Capabilities</i>	
Savita Bhat, Vasudeva Varma and Niranjana Pedanekar	187
<i>Measuring Spurious Correlation in Classification: "Clever Hans" in Translationese</i>	
Angana Borah, Daria Pylypenko, Cristina España-Bonet and Josef van Genabith	196
<i>WIKITIDE: A Wikipedia-Based Timestamped Definition Pairs Dataset</i>	
Hsuvas Borkakoty and Luis Espinosa Anke	207
<i>BERTaporu: Assessing a Genre-Specific Language Model for Portuguese NLP</i>	
Pablo Botton Costa, Matheus Camasmie Pavan, Wesley Ramos Santos, Samuel Caetano Silva and Ivandré Paraboni	217
<i>Comparison of Multilingual Entity Linking Approaches</i>	
Ivelina Bozhinova and Andrey Tagarev	224
<i>Automatic Extraction of the Romanian Academic Word List: Data and Methods</i>	
Ana-Maria Bucur, Andreea Dincă, Madalina Chitez and Roxana Rogobete	234
<i>Stance Prediction from Multimodal Social Media Data</i>	
Lais Carraro Leme Cavalheiro, Matheus Camasmie Pavan and Ivandré Paraboni	242
<i>From Stigma to Support: A Parallel Monolingual Corpus and NLP Approach for Neutralizing Mental Illness Bias</i>	
Mason Choey	249
<i>BB25HLegalSum: Leveraging BM25 and BERT-Based Clustering for the Summarization of Legal Documents</i>	
Leonardo de Andrade and Karin Becker	255
<i>SSSD: Leveraging Pre-trained Models and Semantic Search for Semi-supervised Stance Detection</i>	
André de Sousa and Karin Becker	264
<i>Detecting Text Formality: A Study of Text Classification Approaches</i>	
Daryna Dementieva, Nikolay Babakov and Alexander Panchenko	274
<i>Developing a Multilingual Corpus of Wikipedia Biographies</i>	
Hannah Devinney, Anton Eklund, Igor Ryazanov and Jingwen Cai	285
<i>A Computational Analysis of the Voices of Shakespeare's Characters</i>	
Liviu P. Dinu and Ana Sabina Uban	295

<i>Source Code Plagiarism Detection with Pre-Trained Model Embeddings and Automated Machine Learning</i>	
Fahad Ebrahim and Mike Joy	301
<i>Identifying Semantic Argument Types in Predication and Copredication Contexts: A Zero-Shot Cross-Lingual Approach</i>	
Deniz Ekin Yavas, Laura Kallmeyer, Rainer Osswald, Elisabetta Jezek, Marta Ricchiardi and Long Chen	310
<i>A Review of Research-Based Automatic Text Simplification Tools</i>	
Isabel Espinosa-Zaragoza, José Abreu-Salas, Elena Lloret, Paloma Moreda and Manuel Palomar	321
<i>Vocab-Expander: A System for Creating Domain-Specific Vocabularies Based on Word Embeddings</i>	
Michael Faerber and Nicholas Popovic	331
<i>On the Generalization of Projection-Based Gender Debiasing in Word Embedding</i>	
Elisabetta Fersini, Antonio Candelieri and Lorenzo Pastore	336
<i>Mapping Explicit and Implicit Discourse Relations between the RST-DT and the PDTB 3.0</i>	
Nelson Filipe Costa, Nadia Sheikh and Leila Kosseim	344
<i>Bigfoot in Big Tech: Detecting Out of Domain Conspiracy Theories</i>	
Matthew Fort, Zuoyu Tian, Elizabeth Gabel, Nina Georgiades, Noah Sauer, Daniel Dakota and Sandra Kübler	353
<i>Deep Learning Approaches to Detecting Safeguarding Concerns in Schoolchildren’s Online Conversations</i>	
Emma Franklin and Tharindu Ranasinghe	364
<i>On the Identification and Forecasting of Hate Speech in Inceldom</i>	
Paolo Gajo, Arianna Muti, Katerina Korre, Silvia Bernardini and Alberto Barrón-Cedeño	373
<i>T2KG: Transforming Multimodal Document to Knowledge Graph</i>	
Santiago Galiano, Rafael Muñoz, Yoan Gutiérrez, Andrés Montoyo, Jose Ignacio Abreu and Luis Alfonso Ureña	385
<i>!Translate: When You Cannot Cook Up a Translation, Explain</i>	
Federico Garcea, Margherita Martinelli, Maja Milicević Petrović and Alberto Barrón-Cedeño .	392
<i>An Evaluation of Source Factors in Concatenation-Based Context-Aware Neural Machine Translation</i>	
Harritxu Gete and Thierry Etchegoyhen	399
<i>Lessons Learnt from Linear Text Segmentation: a Fair Comparison of Architectural and Sentence Encoding Strategies for Successful Segmentation</i>	
Iacopo Ghinassi, Lin Wang, Chris Newell and Matthew Purver	408
<i>Student’s t-Distribution: On Measuring the Inter-Rater Reliability When the Observations are Scarce</i>	
Serge Gladkoff, Lifeng Han and Goran Nenadic	419
<i>Data Augmentation for Fake News Detection by Combining Seq2seq and NLI</i>	
Anna Glazkova	429

<i>Exploring Unsupervised Semantic Similarity Methods for Claim Verification in Health Care News Articles</i>	Vishwani Gupta, Astrid Viciano, Holger Wormer and Najmehsadat Mousavinezhad	440
<i>AlphaMWE-Arabic: Arabic Edition of Multilingual Parallel Corpora with Multiword Expression Annotations</i>	najet hadj mohamed, Malak Rassem, Lifeng Han and Goran Nenadic	448
<i>Performance Analysis of Arabic Pre-trained Models on Named Entity Recognition Task</i>	Abdelhalim Hafedh Dahou, Mohamed Amine Cheragui and Ahmed Abdelali	458
<i>Discourse Analysis of Argumentative Essays of English Learners Based on CEFR Level</i>	Blaise Hanel and Leila Kosseim	468
<i>Improving Translation Quality for Low-Resource Inuktitut with Various Preprocessing Techniques</i>	Mathias Hans Erik Stenlund, Mathilde Nanni, Micaella Bruton and Meriem Beloucif	475
<i>Enriched Pre-trained Transformers for Joint Slot Filling and Intent Detection</i>	Momchil Hardalov, Ivan Koychev and Preslav Nakov	480
<i>Unimodal Intermediate Training for Multimodal Meme Sentiment Classification</i>	Muzhaffar Hazman, Susan McKeever and Josephine Griffith	494
<i>Explainable Event Detection with Event Trigger Identification as Rationale Extraction</i>	Hansi Hettiarachchi and Tharindu Ranasinghe	507
<i>Clinical Text Classification to SNOMED CT Codes Using Transformers Trained on Linked Open Medical Ontologies</i>	Anton Hristov, Petar Ivanov, Anna Aksenova, Tsvetan Asamov, Pavlin Gyurov, Todor Primov and Svetla Boytcheva	519
<i>Towards a Consensus Taxonomy for Annotating Errors in Automatically Generated Text</i>	Rudali Huidrom and Anya Belz	527
<i>Uncertainty Quantification of Text Classification in a Multi-Label Setting for Risk-Sensitive Systems</i>	Jinha Hwang, Carol Gudumotu and Benyamin Ahmadnia	541
<i>Pretraining Language- and Domain-Specific BERT on Automatically Translated Text</i>	Tatsuya Ishigaki, Yui Uehara, Goran Topić and Hiroya Takamura	548
<i>Categorising Fine-to-Coarse Grained Misinformation: An Empirical Study of the COVID-19 Infodemic</i>	Ye Jiang, Xingyi Song, Carolina Scarton, Iknor Singh, Ahmet Aker and Kalina Bontcheva	556
<i>Bridging the Gap between Subword and Character Segmentation in Pretrained Language Models</i>	Shun Kiyono, Sho Takase, Shengzhe Li and Toshinori Sato	568
<i>Evaluating Data Augmentation for Medication Identification in Clinical Notes</i>	Jordan Koontz, Maite Oronoz and Alicia Pérez	578
<i>Advancing Topical Text Classification: A Novel Distance-Based Method with Contextual Embeddings</i>	Andriy Kosar, Guy De Pauw and Walter Daelemans	586
<i>Taxonomy-Based Automation of Prior Approval Using Clinical Guidelines</i>	Saranya Krishnamoorthy and Ayush Singh	598

<i>Simultaneous Interpreting as a Noisy Channel: How Much Information Gets Through</i> Maria Kuniilovskaya, Heike Przybyl, Ekaterina Lapshinova-Koltunski and Elke Teich	608
<i>Challenges of GPT-3-Based Conversational Agents for Healthcare</i> Fabian Lechner, Allison Lahnala, Charles Welch and Lucie Flek	619
<i>Noisy Self-Training with Data Augmentations for Offensive and Hate Speech Detection Tasks</i> João Leite, Carolina Scarton and Diego Silva	631
<i>A Practical Survey on Zero-Shot Prompt Design for In-Context Learning</i> Yinheng Li	641
<i>Classifying COVID-19 Vaccine Narratives</i> Yue Li, Carolina Scarton, Xingyi Song and Kalina Bontcheva	648
<i>Sign Language Recognition and Translation: A Multi-Modal Approach Using Computer Vision and Natural Language Processing</i> Jacky Li, Jaren Gerdes, James Gojit, Austin Tao, Samyak Katke, Kate Nguyen and Benyamin Ahmadnia	658
<i>Classification-Aware Neural Topic Model Combined with Interpretable Analysis - for Conflict Classification</i> Tianyu Liang, Yida Mu, Soonho Kim, Darline Kuate, Julie Lang, Rob Vos and Xingyi Song ..	666
<i>Data Augmentation for Fake Reviews Detection</i> Ming Liu and Massimo Poesio	673
<i>Coherent Story Generation with Structured Knowledge</i> Congda Ma, Kotaro Funakoshi, Kiyooki Shirai and Manabu Okumura	681
<i>Studying Common Ground Instantiation Using Audio, Video and Brain Behaviours: The BrainKT Corpus</i> Eliot Maës, Thierry Legou, Leonor Becerra and Philippe Blache	691
<i>Reading between the Lines: Information Extraction from Industry Requirements</i> Ole Magnus Holter and Basil Ell	703
<i>Transformer-Based Language Models for Bulgarian</i> Iva Marinova, Kiril Simov and Petya Osenova	712
<i>Multi-task Ensemble Learning for Fake Reviews Detection and Helpfulness Prediction: A Novel Approach</i> Alimuddin Melleng, Anna Jurek-Loughrey and Deepak P	721
<i>Data Fusion for Better Fake Reviews Detection</i> Alimuddin Melleng, Anna Jurek-Loughrey and Deepak P	730
<i>Dimensions of Quality: Contrasting Stylistic vs. Semantic Features for Modelling Literary Quality in 9,000 Novels</i> Pascale Moreira and Yuri Bizzoni	739
<i>BanglaBait: Semi-Supervised Adversarial Approach for Clickbait Detection on Bangla Clickbait Dataset</i> Md. Motahar Mahtab, Monirul Haque, Mehedi Hasan and Farig Sadeque	748
<i>TreeSwap: Data Augmentation for Machine Translation via Dependency Subtree Swapping</i> Attila Nagy, Dorina Lakatos, Botond Barta and Judit Ács	759

<i>Automatic Assessment Of Spoken English Proficiency Based on Multimodal and Multitask Transformers</i> Kamel Nebhi and György Szaszák	769
<i>Medical Concept Mention Identification in Social Media Posts Using a Small Number of Sample References</i> Vasudevan Nedumpozhimana, Sneha Rautmare, Meegan Gower, Nishtha Jain, Maja Popović, Patricia Buffini and John Kelleher	777
<i>Context-Aware Module Selection in Modular Dialog Systems</i> Jan Nehring, René Marcel Berk and Stefan Hillmann	785
<i>Human Value Detection from Bilingual Sensory Product Reviews</i> Boyu Niu, Céline Manetta and Frédérique Segond	792
<i>Word Sense Disambiguation for Automatic Translation of Medical Dialogues into Pictographs</i> Magali Norré, Rémi Cardon, Vincent Vandeghinste and Thomas François	803
<i>A Research-Based Guide for the Creation and Deployment of a Low-Resource Machine Translation System</i> John E. Ortega and Kenneth Church	813
<i>MQDD: Pre-training of Multimodal Question Duplicity Detection for Software Engineering Domain</i> Jan Pasek, Jakub Sido, Miloslav Konopik and Ondrej Prazak	824
<i>Forming Trees with Treeformers</i> Nilay Patel and Jeffrey Flanigan	836
<i>Evaluating Unsupervised Hierarchical Topic Models Using a Labeled Dataset</i> Judicael Poumay and Ashwin Ittoo	846
<i>HTMOT: Hierarchical Topic Modelling over Time</i> Judicael Poumay and Ashwin Ittoo	854
<i>Multilingual Continual Learning Approaches for Text Classification</i> Karan Praharaj and Irina Matveeva	864
<i>Can Model Fusing Help Transformers in Long Document Classification? An Empirical Study</i> Damith Premasiri, Tharindu Ranasinghe and Ruslan Mitkov	871
<i>Deep Learning Methods for Identification of Multiword Flower and Plant Names</i> Damith Premasiri, Amal Haddad Haddad, Tharindu Ranasinghe and Ruslan Mitkov	879
<i>Improving Aspect-Based Sentiment with End-to-End Semantic Role Labeling Model</i> Pavel Přibán and Ondrej Prazak	888
<i>huPWKP: A Hungarian Text Simplification Corpus</i> Noémi Prótár and Dávid Márk Nemeskey	898
<i>Topic Modeling Using Community Detection on a Word Association Graph</i> Mahfuzur Rahman Chowdhury, Intesur Ahmed, Farig Sadeque and Muhammad Yanhaona	908
<i>Exploring Techniques to Detect and Mitigate Non-Inclusive Language Bias in Marketing Communications Using a Dictionary-Based Approach</i> Bharathi Raja Chakravarthi, Prasanna Kumar Kumaresan, Rahul Ponnusamy, John P. McCrae, Michaela Comerford, Jay Megaro, Deniz Keles and Last Feremenga	918

<i>Does the "Most Sinfully Decadent Cake Ever" Taste Good? Answering Yes/No Questions from Figurative Contexts</i>	
Geetanjali Rakshit and Jeffrey Flanigan	926
<i>Modeling Easiness for Training Transformers with Curriculum Learning</i>	
Leonardo Ranaldi, Giulia Pucci and Fabio Massimo Zanzotto	937
<i>The Dark Side of the Language: Pre-trained Transformers in the DarkNet</i>	
Leonardo Ranaldi, Aria Nourbakhsh, Elena Sofia Ruzzetti, Arianna Patrizi, Dario Onorati, Michele Mastromattei, Francesca Fallucchi and Fabio Massimo Zanzotto	949
<i>PreCog: Exploring the Relation between Memorization and Performance in Pre-trained Language Models</i>	
Leonardo Ranaldi, Elena Sofia Ruzzetti and Fabio Massimo Zanzotto	961
<i>Publish or Hold? Automatic Comment Moderation in Luxembourgish News Articles</i>	
Tharindu Ranasinghe, Alistair Plum, Christoph Purschke and Marcos Zampieri	968
<i>Cross-Lingual Speaker Identification for Indian Languages</i>	
Amaan Rizvi, Anupam Jamatia, Dwijen Rudrapal, Kunal Chakma and Björjn Gambäck	979
<i>'ChemXtract' A System for Extraction of Chemical Events from Patent Documents</i>	
Pattabhi RK Rao and Sobha Lalitha Devi	988
<i>Mind the User! Measures to More Accurately Evaluate the Practical Value of Active Learning Strategies</i>	
Julia Romberg	996
<i>Event Annotation and Detection in Kannada-English Code-Mixed Social Media Data</i>	
Sumukh S, Abhinav Appidi and Manish Shrivastava	1007
<i>Three Approaches to Client Email Topic Classification</i>	
Branislava Šandrih Todorović, Katarina Josipović and Jurij Kodre	1015
<i>Exploring Abstractive Text Summarisation for Podcasts: A Comparative Study of BART and T5 Models</i>	
Parth Saxena and Mo El-Haj	1023
<i>Exploring the Landscape of Natural Language Processing Research</i>	
Tim Schopf, Karim Arabi and Florian Matthes	1034
<i>Efficient Domain Adaptation of Sentence Embeddings Using Adapters</i>	
Tim Schopf, Dennis N. Schneider and Florian Matthes	1046
<i>AspectCSE: Sentence Embeddings for Aspect-Based Semantic Textual Similarity Using Contrastive Learning and Structured Knowledge</i>	
Tim Schopf, Emanuel Gerber, Malte Ostendorff and Florian Matthes	1054
<i>Tackling the Myriads of Collusion Scams on YouTube Comments of Cryptocurrency Videos</i>	
Sadat Shahriar and Arjun Mukherjee	1066
<i>Exploring Deceptive Domain Transfer Strategies: Mitigating the Differences among Deceptive Domains</i>	
Sadat Shahriar, Arjun Mukherjee and Omprakash Gnawali	1076
<i>Party Extraction from Legal Contract Using Contextualized Span Representations of Parties</i>	
Sanjeevan Sivapiran, Charangan Vasantharajan and Uthayasanker Thayasivam	1085

<i>From Fake to Hyperpartisan News Detection Using Domain Adaptation</i>	
Răzvan-Alexandru Smădu, Sebastian-Vasile Echim, Dumitru-Clementin Cercel, Iuliana Marin and Florin Pop	1095
<i>Prompt-Based Approach for Czech Sentiment Analysis</i>	
Jakub Šmíd and Pavel Přibán	1110
<i>Measuring Gender Bias in Natural Language Processing: Incorporating Gender-Neutral Linguistic Forms for Non-Binary Gender Identities in Abusive Speech Detection</i>	
Nasim Sobhani, Kinshuk Sengupta and Sarah Jane Delany	1121
<i>LeSS: A Computationally-Light Lexical Simplifier for Spanish</i>	
Sanja Stajner, Daniel Ibanez and Horacio Saggion	1132
<i>Hindi to Dravidian Language Neural Machine Translation Systems</i>	
Vijay Sundar Ram and Sobha Lalitha Devi	1143
<i>Looking for Traces of Textual Deepfakes in Bulgarian on Social Media</i>	
Irina Temnikova, Iva Marinova, Silvia Gargova, Ruslana Margova and Ivan Koychev	1151
<i>Propaganda Detection in Russian Telegram Posts in the Scope of the Russian Invasion of Ukraine</i>	
Natalia Vanetik, Marina Litvak, Egor Reviakin and Margarita Tiamanova	1162
<i>Auto-Encoding Questions with Retrieval Augmented Decoding for Unsupervised Passage Retrieval and Zero-Shot Question Generation</i>	
Stalin Varanasi, Muhammad Umer Tariq Butt and Guenter Neumann	1171
<i>NoHateBrazil: A Brazilian Portuguese Text Offensiveness Analysis System</i>	
Francielle Vargas, Isabelle Carvalho, Wolfgang Schmeisser-Nieto, Fabrício Benevenuto and Thiago Pardo	1180
<i>Socially Responsible Hate Speech Detection: Can Classifiers Reflect Social Stereotypes?</i>	
Francielle Vargas, Isabelle Carvalho, Ali Hürriyetoglu, Thiago Pardo and Fabrício Benevenuto	1187
<i>Predicting Sentence-Level Factuality of News and Bias of Media Outlets</i>	
Francielle Vargas, Kokil Jaidka, Thiago Pardo and Fabrício Benevenuto	1197
<i>Classification of US Supreme Court Cases Using BERT-Based Techniques</i>	
Shubham Vatsal, Adam Meyers and John E. Ortega	1207
<i>Kāraka-Based Answer Retrieval for Question Answering in Indic Languages</i>	
Devika Verma, Ramprasad S. Joshi, Aiman A. Shivani and Rohan D. Gupta	1216
<i>Comparative Analysis of Named Entity Recognition in the Dungeons and Dragons Domain</i>	
gayashan Weerasundara and Nisansa de Silva	1225
<i>Comparative Analysis of Anomaly Detection Algorithms in Text Data</i>	
Yizhou Xu, Kata Gábor, Jérôme Milleret and Frédérique Segond	1234
<i>Poetry Generation Combining Poetry Theme Labels Representations</i>	
Yingyu Yan, Dongzhen Wen, Liang Yang, Dongyu Zhang and Hongfei LIN	1246
<i>Evaluating Generative Models for Graph-to-Text Generation</i>	
Shuzhou Yuan and Michael Faerber	1256

<i>Microsyntactic Unit Detection Using Word Embedding Models: Experiments on Slavic Languages</i>	
Iuliia Zaitova, Irina Stenger and Tania Avgustinova	1265
<i>Systematic TextRank Optimization in Extractive Summarization</i>	
Morris Zieve, Anthony Gregor, Frederik Juul Stokbaek, Hunter Lewis, Ellis Marie Mendoza and Benyamin Ahmadnia	1274