

13th International Conference on Recent Advances in Natural Language Processing (RANLP 2021)

Deep Learning for Natural Language
Processing Methods and Applications

Online
1 – 3 September 2021

Volume 1 of 3

Editors:

**Galia Angelova
Maria Kunilovskaya**

**Ruslan Mitkov
Ivelina Nikolova-Koleva**

ISBN: 978-1-7138-4095-4

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2021) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2022)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006

Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>BPoMP: The Benchmark of Poetic Minimal Pairs – Limericks, Rhyme, and Narrative Coherence</i> Almas Abdibayev, Allen Riddell and Daniel Rockmore	1
<i>Ontology Population Reusing Resources for Dialogue Intent Detection: Generic and Multilingual Approach</i> Cristina Aceta, Izaskun Fernández and Aitor Soroa	10
<i>Efficient Multilingual Text Classification for Indian Languages</i> Salil Aggarwal, Sourav Kumar and Radhika Mamidi	19
<i>Domain Adaptation for Hindi-Telugu Machine Translation Using Domain Specific Back Translation</i> Hema Ala, Vandan Mujadia and Dipti Sharma	26
<i>ArabGlossBERT: Fine-Tuning BERT on Context-Gloss Pairs for WSD</i> Moustafa Al-Hajj and Mustafa Jarrar	35
<i>English-Arabic Cross-language Plagiarism Detection</i> Naif Alotaibi and Mike Joy	44
<i>Towards a Better Understanding of Noise in Natural Language Processing</i> Khetam Al Sharou, Zhenhao Li and Lucia Specia	53
<i>Comparing Supervised Machine Learning Techniques for Genre Analysis in Software Engineering Research Articles</i> Felipe Araújo de Britto, Thiago Castro Ferreira, Leonardo Pereira Nunes and Fernando Silva Parreiras	63
<i>Enriching the Transformer with Linguistic Factors for Low-Resource Machine Translation</i> Jordi Armengol-Estapé, Marta R. Costa-jussà and Carlos Escolano	73
<i>A Multi-Pass Sieve Coreference Resolution for Indonesian</i> Valentina Kania Prameswara Artari, Rahmad Mahendra, Meganingrum Arista Jiwanggi, Adityo Anggraito and Indra Budi	79
<i>Solving SCAN Tasks with Data Augmentation and Input Embeddings</i> Michal Auersperger and Pavel Pecina	86
<i>PyEuroVoc: A Tool for Multilingual Legal Document Classification with EuroVoc Descriptors</i> Andrei-Marius Avram, Vasile Pais and Dan Ioan Tufis	92
<i>TEASER: Towards Efficient Aspect-based SENTiment Analysis and Recognition</i> Vaibhav Bajaj, Kartikey Pant, Ishan Upadhyay, Srinath Nair and Radhika Mamidi	102
<i>Interactive Learning Approach for Arabic Target-Based Sentiment Analysis</i> Husamelddin Balla, Marisa Llorens Salvador and Sarah Jane Delany	111
<i>Litescale: A Lightweight Tool for Best-worst Scaling Annotation</i> Valerio Basile and Christian Cagnazzo	121
<i>Probabilistic Ensembles of Zero- and Few-Shot Learning Models for Emotion Classification</i> Angelo Basile, Guillermo Pérez-Torró and Marc Franco-Salvador	128

<i>Cross-Lingual Wolastoqey-English Definition Modelling</i>	
Diego Bear and Paul Cook	138
<i>Neural Network-Based Generation of Sport Summaries: A Preliminary Study</i>	
David Stéphane Belemkoabga, Aurélien Bossard, Abdallah Essa, Christophe Rodrigues and Kévin Sylla	147
<i>Split-and-Rephrase in a Cross-Lingual Manner: A Complete Pipeline</i>	
Paulo Berlanga Neto and Evandro Eduardo Seron Ruiz	155
<i>On the Contribution of Per-ICD Attention Mechanisms to Classify Health Records in Languages with Fewer Resources than English</i>	
Alberto Blanco, Sonja Remmer, Alicia Pérez, Hercules Dalianis and Arantza Casillas	165
<i>Can the Transformer Be Used as a Drop-in Replacement for RNNs in Text-Generating GANs?</i>	
Kevin Blin and Andrei Kucharyv	173
<i>Predicting the Factuality of Reporting of News Media Using Observations about User Attention in Their YouTube Channels</i>	
Krasimira Bozhanova, Yoan Dinkov, Ivan Koychev, Maria Castaldo, Tommaso Venturini and Preslav Nakov	182
<i>OCR Processing of Swedish Historical Newspapers Using Deep Hybrid CNN–LSTM Networks</i>	
Molly Brandt Skelbye and Dana Dannélls	190
<i>A Psychologically Informed Part-of-Speech Analysis of Depression in Social Media</i>	
Ana-Maria Bucur, Ioana R. Podina and Liviu P. Dinu	199
<i>InFoBERT: Zero-Shot Approach to Natural Language Understanding Using Contextualized Word Embedding</i>	
Pavel Burnyshev, Andrey Bout, Valentin Malykh and Irina Piontkovskaya	208
<i>Active Learning for Assisted Corpus Construction: A Case Study in Knowledge Discovery from Biomedical Text</i>	
Hian Cañizares-Díaz, Alejandro Piad-Morffis, Suilan Estevez-Velarde, Yoan Gutiérrez, Yudivián Almeida Cruz, Andres Montoyo and Rafael Muñoz-Guillena	216
<i>Unsupervised Text Style Transfer with Content Embeddings</i>	
Keith Carlson, Allen Riddell and Daniel Rockmore	226
<i>Evaluating Recognizing Question Entailment Methods for a Portuguese Community Question-Answering System about Diabetes Mellitus</i>	
Thiago Castro Ferreira, João Victor de Pinho Costa, Isabela Rigotto, Vitoria Portella, Gabriel Frota, Ana Luisa A. R. Guimarães, Adalberto Penna, Isabela Lee, Tayane A. Soares, Sophia Rolim, Rossana Cunha, Celso França, Ariel Santos, Rivaney F. Oliveira, Abisague Langbehn, Daniel Hasan Dalip, Marcos André Gonçalves, Rodrigo Bastos Fóscolo and Adriana Pagano	234
<i>On the Usability of Transformers-based Models for a French Question-Answering Task</i>	
Oralie Cattan, Christophe Servan and Sophie Rosset	244
<i>Classification of Code-Mixed Text Using Capsule Networks</i>	
Shanaka Chathuranga and Surangika Ranathunga	256
<i>Character-based Thai Word Segmentation with Multiple Attentions</i>	
Thodsaporn Chay-intr, Hidetaka Kamigaito and Manabu Okumura	264

<i>Are Language-Agnostic Sentence Representations Actually Language-Agnostic?</i> Yu Chen and Tania Avgustinova	274
<i>Investigating Dominant Word Order on Universal Dependencies with Graph Rewriting</i> Hee-Soo Choi, Bruno Guillaume, Karèn Fort and Guy Perrier	281
<i>RED: A Novel Dataset for Romanian Emotion Detection from Tweets</i> Alexandra Ciobotaru and Liviu P. Dinu	291
<i>Assessing the Eligibility of Backtranslated Samples Based on Semantic Similarity for the Paraphrase Identification Task</i> Jean-Philippe Corbeil and Hadi Abdi Ghavidel	301
<i>Fine-tuning Neural Language Models for Multidimensional Opinion Mining of English-Maltese Social Data</i> Keith Cortis, Kanishk Verma and Brian Davis	309
<i>Towards an Etymological Map of Romanian</i> Alina Maria Cristea, Anca Dinu, Liviu P. Dinu, Simona Georgescu, Ana Sabina Uban and Laurentiu Zoicas	315
<i>A Syntax-Aware Edit-based System for Text Simplification</i> Oscar M. Cumbicus-Pineda, Itziar Gonzalez-Dios and Aitor Soroa	324
<i>On Generating Fact-Infused Question Variations</i> Arthur Deschamps, Sujatha Das Gollapalli and See-Kiong Ng	335
<i>Event Prominence Extraction Combining a Knowledge-Based Syntactic Parser and a BERT Classifier for Dutch</i> Thierry Desot, Orphee De Clercq and Veronique Hoste	346
<i>Automatic Detection and Classification of Mental Illnesses from General Social Media Texts</i> Anca Dinu and Andreea-Codrina Moldovan	358
<i>A Pre-trained Transformer and CNN Model with Joint Language ID and Part-of-Speech Tagging for Code-Mixed Social-Media Text</i> Suman Dowlagar and Radhika Mamidi	367
<i>Tracing Source Language Interference in Translation with Graph-Isomorphism Measures</i> Koel Dutta Chowdhury, Cristina España-Bonet and Josef van Genabith	375
<i>Decoupled Transformer for Scalable Inference in Open-domain Question Answering</i> Haytham Elfdaeel and Stanislav Peshterliev	386
<i>Towards Task-Agnostic Privacy- and Utility-Preserving Models</i> Yaroslav Emelyanov	394
<i>Knowledge Discovery in COVID-19 Research Literature</i> Ernesto L. Estevanell-Valladares, Suilan Estevez-Velarde, Alejandro Piad-Morffis, Yoan Gutierrez, Andres Montoyo, Rafael Muñoz and Yudiivián Almeida Cruz	402
<i>Online Learning over Time in Adaptive Neural Machine Translation</i> Thierry Etchegoyhen, David Ponce, Harritxu Gete and Victor Ruiz	411

<i>Improving Character-Aware Neural Language Model by Warming up Character Encoder under Skip-gram Architecture</i>	
Yukun Feng, Chenlong Hu, Hidetaka Kamigaito, Hiroya Takamura and Manabu Okumura	421
<i>Interpretable Identification of Cybersecurity Vulnerabilities from News Articles</i>	
Pierre Frode de la Foret, Stefan Ruseti, Cristian Sandescu, Mihai Dascalu and Sebastien Travadel	428
<i>Cross-lingual Offensive Language Identification for Low Resource Languages: The Case of Marathi</i>	
Saurabh Sampatrao Gaikwad, Tharindu Ranasinghe, Marcos Zampieri and Christopher Homan	437
<i>Relying on Discourse Analysis to Answer Complex Questions by Neural Machine Reading Comprehension</i>	
Boris Galitsky, Dmitry Ilvovsky and Elizaveta Goncharova	444
<i>A Dynamic Head Importance Computation Mechanism for Neural Machine Translation</i>	
Akshay Goindani and Manish Shrivastava	454
<i>Syntax and Themes: How Context Free Grammar Rules and Semantic Word Association Influence Book Success</i>	
Henry Gorelick, Biddut Sarker Bijoy, Syeda Jannatus Saba, Sudipta Kar, Md Saiful Islam and Mohammad Ruhul Amin	463
<i>SocialVisTUM: An Interactive Visualization Toolkit for Correlated Neural Topic Models on Social Media Opinion Mining</i>	
Gerhard Hagerer, Martin Kirchhoff, Hannah Danner, Robert Pesch, Mainak Ghosh, Archishman Roy, Jiaxi Zhao and Georg Groh	475
<i>Apples to Apples: A Systematic Evaluation of Topic Models</i>	
Ismail Harrando, Pasquale Lisena and Raphael Troncy	483
<i>Claim Verification Using a Multi-GAN Based Model</i>	
Amartya Hatua, Arjun Mukherjee and Rakesh Verma	494
<i>Semi-Supervised and Unsupervised Sense Annotation via Translations</i>	
Bradley Hauer, Grzegorz Kondrak, Yixing Luan, Arnob Mallik and Lili Mou	504
<i>Personality Predictive Lexical Cues and Their Correlations</i>	
Xiaoli He and Gerard de Melo	514
<i>Evaluation Datasets for Cross-lingual Semantic Textual Similarity</i>	
Tomáš Hercig and Pavel Kral	524
<i>Relation Extraction Using Multiple Pre-Training Models in Biomedical Domain</i>	
Satoshi Hiai, Kazutaka Shimada, Taiki Watanabe, Akiva Miura and Tomoya Iwakura	530
<i>Discussion Structure Prediction Based on a Two-step Method</i>	
Takumi Himeno and Kazutaka Shimada	538
<i>On the Usefulness of Personality Traits in Opinion-oriented Tasks</i>	
Marjan Hosseinia, Eduard Dragut, Dainis Boumber and Arjun Mukherjee	547

<i>Application of Deep Learning Methods to SNOMED CT Encoding of Clinical Texts: From Data Collection to Extreme Multi-Label Text-Based Classification</i>	
Anton Hristov, Aleksandar Tahchiev, Hristo Papazov, Nikola Tulechki, Todor Primov and Svetla Boytcheva	557
<i>Syntax Matters! Syntax-Controlled in Text Style Transfer</i>	
Zhiqiang Hu, Roy Ka-Wei Lee and Charu C. Aggarwal	566
<i>Transfer Learning for Czech Historical Named Entity Recognition</i>	
Helena Hubková and Pavel Kral	576
<i>Personality Trait Identification Using the Russian Feature Extraction Toolkit</i>	
James R. Hull, Valerie Novak, C. Anton Rytting, Paul Rodrigues, Victor M. Frank and Matthew Swahn	583
<i>Semi-Supervised Learning Based on Auto-generated Lexicon Using XAI in Sentiment Analysis</i>	
Hohyun Hwang and Younghoon Lee	593
<i>Multiple Teacher Distillation for Robust and Greener Models</i>	
Artur Ilichev, Nikita Sorokin, Irina Piontkovskaya and Valentin Malykh	601
<i>BERT Embeddings for Automatic Readability Assessment</i>	
Joseph Marvin Imperial	611
<i>Semantic-Based Opinion Summarization</i>	
Marcio Inácio and Thiago Pardo	619
<i>Using Collaborative Filtering to Model Argument Selection</i>	
Sagar Indurkha	629
<i>Domain-Specific Japanese ELECTRA Model Using a Small Corpus</i>	
Youki Itoh and Hiroyuki Shinnou	640
<i>BERT-PersNER: A New Model for Persian Named Entity Recognition</i>	
Farane Jalali Farahani and Gholamreza Ghassem-Sani	647
<i>Cross-lingual Fine-tuning for Abstractive Arabic Text Summarization</i>	
Mram Kahla, Zijian Győző Yang and Attila Novák	655
<i>Behavior of Modern Pre-trained Language Models Using the Example of Probing Tasks</i>	
Ekaterina Kalyaeva, Oleg Durandin and Alexey Malafeev	664
<i>Towards Quantifying Magnitude of Political Bias in News Articles Using a Novel Annotation Schema</i>	
Lalitha Kameswari and Radhika Mamidi	671
<i>Application of Mix-Up Method in Document Classification Task Using BERT</i>	
Naoki Kikuta and Hiroyuki Shinnou	679
<i>Translation Memory Retrieval Using Lucene</i>	
Kwang-hyok Kim, Myong-ho Cho, Chol-ho Ryang, Ju-song Im, Song-yong Cho and Yong-jun Han	684
<i>Now, It's Personal : The Need for Personalized Word Sense Disambiguation</i>	
Milton King and Paul Cook	692

<i>Multilingual Image Corpus: Annotation Protocol</i>	
Svetla Koeva	701
<i>ELERRANT: Automatic Grammatical Error Type Classification for Greek</i>	
Katerina Korre, Marita Chatzipanagiotou and John Pavlopoulos	708
<i>Neural Machine Translation for Sinhala-English Code-Mixed Text</i>	
Archchana Kugathanan and Sagara Sumathipala	718
<i>Multilingual Multi-Domain NMT for Indian Languages</i>	
Sourav Kumar, Salil Aggarwal and Dipti Sharma	727
<i>Fiction in Russian Translation: A Translationese Study</i>	
Maria Kunilovskaya, Ekaterina Lapshinova-Koltunski and Ruslan Mitkov	734
<i>Corpus Creation and Language Identification in Low-Resource Code-Mixed Telugu-English Text</i>	
Siva Subrahmanyam Varma Kusampudi, Anudeep Chaluvadi and Radhika Mamidi	744
<i>Sentiment Analysis in Code-Mixed Telugu-English Text with Unsupervised Data Normalization</i>	
Siva Subrahmanyam Varma Kusampudi, Preetham Sathineni and Radhika Mamidi	753
<i>From Constituency to UD-Style Dependency: Building the First Conversion Tool of Turkish</i>	
Aslı Kuzgun, Oğuz Kerem Yıldız, Neslihan Cesur, Büşra Marşan, Arife Betül Yenice, Ezgi Sanyar, Oguzhan Kuyrukçu, Bilge Nas Arıcan and Olcay Taner Yıldız	761
<i>Making Your Tweets More Fancy: Emoji Insertion to Texts</i>	
Jingun Kwon, Naoki Kobayashi, Hidetaka Kamigaito, Hiroya Takamura and Manabu Okumura	770
<i>Addressing Slot-Value Changes in Task-oriented Dialogue Systems through Dialogue Domain Adaptation</i>	
Tiziano Labruna and Bernardo Magnini	780
<i>Developing a Clinical Language Model for Swedish: Continued Pretraining of Generic BERT with In-Domain Data</i>	
Anastasios Lamproudis, Aron Henriksson and Hercules Dalianis	790
<i>Text Retrieval for Language Learners: Graded Vocabulary vs. Open Learner Model</i>	
John Lee and Chak Yan Yeung	798
<i>Transforming Multi-Conditioned Generation from Meaning Representation</i>	
Joosung Lee	805
<i>Frustration Level Annotation in Latvian Tweets with Non-Lexical Means of Expression</i>	
Viktorija Leonova and Janis Zuters	814
<i>System Combination for Grammatical Error Correction Based on Integer Programming</i>	
Ruixi Lin and Hwee Tou Ng	824
<i>Multilingual Learning for Mild Cognitive Impairment Screening from a Clinical Speech Task</i>	
Hali Lindsay, Philipp Müller, Insa Kröger, Johannes Tröger, Nicklas Linz, Alexandra König, Radia Zeghari, Frans RJ Verhey and Inez HGB Ramakers	830
<i>Naturalness Evaluation of Natural Language Generation in Task-oriented Dialogues Using BERT</i>	
Ye Liu, Wolfgang Maier, Wolfgang Minker and Stefan Ultes	839

<i>Towards the Application of Calibrated Transformers to the Unsupervised Estimation of Question Difficulty from Text</i>	
Ekaterina Loginova, Luca Benedetto, Dries Benoit and Paolo Cremonesi	846
<i>GeSERA: General-domain Summary Evaluation by Relevance Analysis</i>	
Jessica López Espejel, Gaël de Chalendar, Jorge Garcia Flores, Thierry Charnois and Ivan Vladimir Meza Ruiz	856
<i>On the Interaction between Annotation Quality and Classifier Performance in Abusive Language Detection</i>	
Holly Lopez Long, Alexandra O’Neil and Sandra Kübler	868
<i>NEREL: A Russian Dataset with Nested Named Entities, Relations and Events</i>	
Natalia Loukachevitch, Ekaterina Artemova, Tatiana Batura, Pavel Braslavski, Iliia Denisov, Vladimir Ivanov, Suresh Manandhar, Alexander Pugachev and Elena Tutubalina	876
<i>Active Learning for Interactive Relation Extraction in a French Newspaper’s Articles</i>	
Cyrielle Mallart, Michel Le Nouy, Guillaume Gravier and Pascale Sébillot	886
<i>ROFF - A Romanian Twitter Dataset for Offensive Language</i>	
Mihai Manolescu and Çağrı Çöltekin	895
<i>Monitoring Fact Preservation, Grammatical Consistency and Ethical Behavior of Abstractive Summarization Neural Models</i>	
Iva Marinova, Yolina Petrova, Milena Slavcheva, Petya Osenova, Ivaylo Radev and Kiril Simov	901
<i>Cultural Topic Modelling over Novel Wikipedia Corpora for South-Slavic Languages</i>	
Filip Markoski, Elena Markoska, Nikola Ljubešić, Eftim Zdravevski and Ljupco Kocarev	910
<i>Discovery of Multiword Expressions with Loanwords and Their Equivalents in the Persian Language</i>	
Katarzyna Marszałek-Kowalewska	918
<i>The Impact of Text Normalization on Multiword Expressions Discovery in Persian</i>	
Katarzyna Marszałek-Kowalewska	929
<i>Improving Neural Language Processing with Named Entities</i>	
Kyومoto Matsushita, Takuya Makino and Tomoya Iwakura	940
<i>TREMoLo-Tweets: A Multi-Label Corpus of French Tweets for Language Register Characterization</i>	
Jade Mekki, Gwénolé Lecorvé, Delphine Battistelli and Nicolas Béchet	950
<i>Ranking Online Reviews Based on Their Helpfulness: An Unsupervised Approach</i>	
Alimuddin Melleng, Anna Jurek-Loughrey and Deepak P.	959
<i>incom.py 2.0 - Calculating Linguistic Distances and Asymmetries in Auditory Perception of Closely Related Languages</i>	
Marius Mosbach, Irina Stenger, Tania Avgustinova, Bernd Möbius and Dietrich Klakow	968
<i>Not All Linearizations Are Equally Data-Hungry in Sequence Labeling Parsing</i>	
Alberto Muñoz-Ortiz, Michalina Strzyz and David Vilares	978
<i>Pre-training a BERT with Curriculum Learning by Increasing Block-Size of Input Text</i>	
Koichi Nagatsuka, Clifford Broni-Bediako and Masayasu Atsumi	989
<i>COVID-19 in Bulgarian Social Media: Factuality, Harmfulness, Propaganda, and Framing</i>	
Preslav Nakov, Firoj Alam, Shaden Shaar, Giovanni Da San Martino and Yifan Zhang	997

<i>A Second Pandemic? Analysis of Fake News about COVID-19 Vaccines in Qatar</i> Preslav Nakov, Firoj Alam, Shaden Shaar, Giovanni Da San Martino and Yifan Zhang	1010
<i>A Hierarchical Entity Graph Convolutional Network for Relation Extraction across Documents</i> Tapas Nayak and Hwee Tou Ng	1022
<i>Improving Distantly Supervised Relation Extraction with Self-Ensemble Noise Filtering</i> Tapas Nayak, Navonil Majumder and Soujanya Poria	1031
<i>Learning Entity-Likeness with Multiple Approximate Matches for Biomedical NER</i> An Nguyen Le, Hajime Morita and Tomoya Iwakura	1040
<i>Extending a Text-to-Pictograph System to French and to Arasaac</i> Magali Norré, Vincent Vandeghinste, Pierrette Bouillon and Thomas François	1050
<i>Transfer-based Enrichment of a Hungarian Named Entity Dataset</i> Attila Novák and Borbála Novák	1060
<i>One Size Does Not Fit All: Finding the Optimal Subword Sizes for FastText Models across Languages</i> Vít Novotný, Eniafe Festus Ayetiran, Dalibor Bačovský, Dávid Lupták, Michal Štefánik and Petr Sojka	1068
<i>CLexIS2: A New Corpus for Complex Word Identification Research in Computing Studies</i> Jenny A. Ortiz Zambrano and Arturo Montejo-Ráez	1075
<i>Towards Precise Lexicon Integration in Neural Machine Translation</i> Ogün Öz and Maria Sukhareva	1084
<i>OffendES: A New Corpus in Spanish for Offensive Language Research</i> Flor Miriam Plaza-del-Arco, Arturo Montejo-Ráez, L. Alfonso Ureña-López and María-Teresa Martín-Valdivia	1096
<i>On Machine Translation of User Reviews</i> Maja Popović, Alberto Poncelas, Marija Brkic and Andy Way	1109
<i>Multilingual Coreference Resolution with Harmonized Annotations</i> Ondřej Pražák, Miloslav Konopík and Jakub Sido	1119
<i>Predicting Informativeness of Semantic Triples</i> Judita Preiss	1124
<i>Unknown Intent Detection Using Multi-Objective Optimization on Deep Learning Classifiers</i> Perna Prem, Zishan Ahmad, Asif Ekbal, Shubhashis Sengupta, Sakshi C. Jain and Roshni Ramnani	1130
<i>Are the Multilingual Models Better? Improving Czech Sentiment with Transformers</i> Pavel Přibáň and Josef Steinberger	1138
<i>Metric Learning in Multilingual Sentence Similarity Measurement for Document Alignment</i> Charith Rajitha, Lakmali Piyarathna, Dilan Sachintha and Surangika Ranathunga	1150
<i>Multi-label Diagnosis Classification of Swedish Discharge Summaries – ICD-10 Code Assignment Using KB-BERT</i> Sonja Remmer, Anastasios Lamproudis and Hercules Dalianis	1158

<i>Siamese Networks for Inference in Malayalam Language Texts</i>	
Sara Renjit and Sumam Mary Idicula	1167
<i>A Call for Clarity in Contemporary Authorship Attribution Evaluation</i>	
Allen Riddell, Haining Wang and Patrick Juola	1174
<i>Varieties of Plain Language</i>	
Allen Riddell and Yohei Igarashi	1180
<i>Word Discriminations for Vocabulary Inventory Prediction</i>	
Frankie Robertson	1188
<i>FrenLyS: A Tool for the Automatic Simplification of French General Language Texts</i>	
Eva Rolin, Quentin Langlois, Patrick Watrin and Thomas François	1196
<i>Spelling Correction for Russian: A Comparative Study of Datasets and Methods</i>	
Alla Rozovskaya	1206
<i>Sentiment-Aware Measure (SAM) for Evaluating Sentiment Transfer by Machine Translation Systems</i>	
Hadeel Saadany, Constantin Orăsan, Emad Mohamed and Ashraf Tantavy	1217
<i>Multilingual Epidemic Event Extraction : From Simple Classification Methods to Open Information Extraction (OIE) and Ontology</i>	
Sihem Sahnoun and Gaël Lejeune	1227
<i>Exploiting Domain-Specific Knowledge for Judgment Prediction Is No Panacea</i>	
Olivier Salaün, Philippe Langlais and Karim Benyekhlef	1234
<i>Masking and Transformer-based Models for Hyperpartisanship Detection in News</i>	
Javier Sánchez-Junquera, Paolo Rosso, Manuel Montes-y-Gómez and Simone Paolo Ponzetto	1244
<i>Serbian NER&Beyond: The Archaic and the Modern Intertwined</i>	
Branislava Šandrih Todorović, Cvetana Krstev, Ranka Stanković and Milica Ikonić Nešić	1252
<i>A Semi-Supervised Approach to Detect Toxic Comments</i>	
Ghivvago Damas Saraiva, Rafael Anchiêta, Francisco Assis Ricarte Neto and Raimundo Moura	1261
<i>Graph-based Argument Quality Assessment</i>	
Ekaterina Saveleva, Volha Petukhova, Marius Mosbach and Dietrich Klakow	1268
<i>A Hybrid Approach of Opinion Mining and Comparative Linguistic Analysis of Restaurant Reviews</i>	
Salim Sazed	1281
<i>A Lexicon for Profane and Obscene Text Identification in Bengali</i>	
Salim Sazed	1289
<i>A Case Study of Deep Learning-Based Multi-Modal Methods for Labeling the Presence of Questionable Content in Movie Trailers</i>	
Mahsa Shafaei, Christos Smailis, Ioannis Kakadiaris and Thamar Solorio	1297
<i>A Domain-Independent Holistic Approach to Deception Detection</i>	
Sadat Shahriar, Arjun Mukherjee and Omprakash Gnawali	1308
<i>Towards Domain-Generalizable Paraphrase Identification by Avoiding the Shortcut Learning</i>	
Xin Shen and Wai Lam	1318

<i>Czert – Czech BERT-like Model for Language Representation</i>	
Jakub Sido, Ondřej Pražák, Pavel Přibáň, Jan Pašek, Michal Seják and Miloslav Konopík	1326
<i>Exploring German Multi-Level Text Simplification</i>	
Nicolas Spring, Annette Rios and Sarah Ebling	1339
<i>Exploring Reliability of Gold Labels for Emotion Detection in Twitter</i>	
Sanja Stajner	1350
<i>How to Obtain Reliable Labels for MBTI Classification from Texts?</i>	
Sanja Stajner and Seren Yenikent	1360
<i>Watching a Language Model Learning Chess</i>	
Andreas Stöckl	1369
<i>Tackling Multilinguality and Internationality in Fake News</i>	
Andrey Tagarev, Krasimira Bozhanova, Ivelina Nikolova-Koleva and Ivan Ivanov	1380
<i>Learning and Evaluating Chinese Idiom Embeddings</i>	
Minghuan Tan and Jing Jiang	1387
<i>Does BERT Understand Idioms? A Probing-Based Empirical Study of BERT Encodings of Idioms</i>	
Minghuan Tan and Jing Jiang	1397
<i>An Empirical Analysis of Topic Models: Uncovering the Relationships between Hyperparameters, Document Length and Performance Measures</i>	
Silvia Terragni and Elisabetta Fersini	1408
<i>TR-SEQ: Named Entity Recognition Dataset for Turkish Search Engine Queries</i>	
Berkay Topçu and İlknur Durgar El-Kahlout	1417
<i>Opinion Prediction with User Fingerprinting</i>	
Kishore Tumarada, Yifan Zhang, Fan Yang, Eduard Dragut, Omprakash Gnawali and Arjun Mukherjee	1423
<i>Can Multilingual Transformers Fight the COVID-19 Infodemic?</i>	
Lasitha Uyangodage, Tharindu Ranasinghe and Hansi Hettiarachchi	1432
<i>Contextual-Lexicon Approach for Abusive Language Detection</i>	
Francielle Vargas, Fabiana Rodrigues de Góes, Isabelle Carvalho, Fabrício Benevenuto and Thiago Pardo	1438
<i>Comparative Analysis of Fine-tuned Deep Learning Language Models for ICD-10 Classification Task for Bulgarian Language</i>	
Boris Velichkov, Sylvia Vassileva, Simeon Gerginov, Boris Kraychev, Ivaylo Ivanov, Philip Ivanov, Ivan Koychev and Svetla Boytcheva	1448
<i>Mistake Captioning: A Machine Learning Approach for Detecting Mistakes and Generating Instructive Feedback</i>	
Anton Vinogradov, Andrew Miles Byrd and Brent Harrison	1455
<i>A Novel Machine Learning Based Approach for Post-OCR Error Detection</i>	
Shafqat Mumtaz Virk, Dana Dannélls and Azam Sheikh Muhammad	1463
<i>A Data-Driven Semi-Automatic Framenet Development Methodology</i>	
Shafqat Mumtaz Virk, Dana Dannélls, Lars Borin and Markus Forsberg	1471

<i>A Deep Learning System for Automatic Extraction of Typological Linguistic Information from Descriptive Grammars</i>	
Shafqat Mumtaz Virk, Daniel Foster, Azam Sheikh Muhammad and Raheela Saleem	1480
<i>Recognizing and Splitting Conditional Sentences for Automation of Business Processes Management</i>	
Ngoc Phuoc An Vo, Irene Manotas, Octavian Popescu, Algimantas Černiauskas and Vadim Sheinin	1490
<i>“Don’t discuss”: Investigating Semantic and Argumentative Features for Supervised Propagandist Message Detection and Classification</i>	
Vorakit Vorakitphan, Elena Cabrio and Serena Villata	1498
<i>ComboNER: A Lightweight All-In-One POS Tagger, Dependency Parser and NER</i>	
Aleksander Wawer	1508
<i>Investigating Annotator Bias in Abusive Language Datasets</i>	
Maximilian Wich, Christian Widmer, Gerhard Hagerer and Georg Groh	1515
<i>Rules Ruling Neural Networks - Neural vs. Rule-Based Grammar Checking for a Low Resource Language</i>	
Linda Wiecheteck, Flammie Pirinen, Mika Hämäläinen and Chiara Argese	1526
<i>Transformer with Syntactic Position Encoding for Machine Translation</i>	
Yikuan Xie, Wenyong Wang, Mingqian Du and Qing He	1536
<i>Towards Sentiment Analysis of Tobacco Products’ Usage in Social Media</i>	
Venkata Himakar Yanamandra, Kartikey Pant and Radhika Mamidi	1545
<i>Improving Evidence Retrieval with Claim-Evidence Entailment</i>	
Fan Yang, Eduard Dragut and Arjun Mukherjee	1553
<i>Sentence Structure and Word Relationship Modeling for Emphasis Selection</i>	
Haoran Yang and Wai Lam	1559
<i>Utterance Position-Aware Dialogue Act Recognition</i>	
Yuki Yano, Akihiro Tamura, Takashi Ninomiya and Hiroaki Obayashi	1567
<i>Tell Me What You Read: Automatic Expertise-Based Annotator Assignment for Text Annotation in Expert Domains</i>	
Hiyori Yoshikawa, Tomoya Iwakura, Kimi Kaneko, Hiroaki Yoshida, Yasutaka Kumano, Kazutaka Shimada, Rafal Rzepka and Patrycja Swieczkowska	1575
<i>Abstractive Document Summarization with Word Embedding Reconstruction</i>	
Jingyi You, Chenlong Hu, Hidetaka Kamigaito, Hiroya Takamura and Manabu Okumura	1586
<i>Interpretable Propaganda Detection in News Articles</i>	
Seunghak Yu, Giovanni Da San Martino, Mitra Mohtarami, James Glass and Preslav Nakov .	1597
<i>Generic Mechanism for Reducing Repetitions in Encoder-Decoder Models</i>	
Ying Zhang, Hidetaka Kamigaito, Tatsuya Aoki, Hiroya Takamura and Manabu Okumura . . .	1606
<i>Knowledge Distillation with BERT for Image Tag-Based Privacy Prediction</i>	
Chenye Zhao and Cornelia Caragea	1616
<i>Delexicalized Cross-lingual Dependency Parsing for Xibe</i>	
He Zhou and Sandra Kübler	1626

AutoChart: A Dataset for Chart-to-Text Generation Task
Jiawen Zhu, Jinye Ran, Roy Ka-Wei Lee, Zhi Li and Kenny Choo 1636

A Comparative Study on Abstractive and Extractive Approaches in Summarization of European Legislation Documents
Valentin Zmiycharov, Milen Chechev, Gergana Lazarova, Todor Tsonkov and Ivan Koychev . 1645

Not All Comments Are Equal: Insights into Comment Moderation from a Topic-Aware Model
Elaine Zosa, Ravi Shekhar, Mladen Karan and Matthew Purver 1652