

ICON 2020

**17th International Conference on Natural Language
Processing**

Proceedings of the Conference

December 18 - 21, 2020
Indian Institute of Technology Patna, India

©2020 NLP Association of India (NLP AI)

Preface

Research in Natural Language Processing (NLP) has taken a noticeable leap in recent years. The tremendous growth of information on the web and its easy access has stimulated a large interest in the field. India, with multiple languages and continuous growth of Indian language content on the web, makes a fertile ground for NLP research. Moreover, the industry is keenly interested in obtaining NLP technology for mass use. Internet search companies are increasingly aware of the large market for processing languages other than English. For example, search capability is needed for content in Indian and other languages. There is also a need for searching content in multiple languages, and making the retrieved documents available in the language of the user. As a result, a strong need is being felt for machine translation to handle this large instantaneous use. Information Extraction, Question Answering Systems, and Sentiment Analysis are also showing up as other business opportunities.

These needs have resulted in two welcome trends. First, there is a much wider student interest in getting into NLP at both postgraduate and undergraduate levels. Many students interested in computing technology are getting interested in natural language technology, and those interested in pursuing computing research are joining NLP research. Second, the research community in academic institutions and government funding agencies in India have joined hands to launch consortia projects to develop NLP products. Each consortium project is a multi-institutional endeavour working with a common software framework, common language standards, and common technology engines for all the different languages covered in the consortium. As a result, it has already led to the development of basic tools for multiple languages that are interoperable for machine translation, cross-lingual search, handwriting recognition, and OCR.

In this backdrop of increased student interest, greater funding, and most importantly, common standards and interoperable tools, there has been a spurt in research in NLP on Indian languages whose effects we have just begun to see. A great number of submissions reflecting good research is a heartening matter. There is an increasing realization to take advantage of features common to Indian languages in machine learning. It is a delight to see that such features are not just specific to Indian languages but to a large number of languages of the world, hitherto ignored. The insights so gained are furthering our linguistic understanding and will help in technology development for hopefully all languages of the world.

For machine learning and other purposes, linguistically annotated corpora using the common standards have become available for multiple Indian languages. They have been used for the development of basic technologies for several languages. A larger set of corpora are expected to be prepared in the near future.

These conference proceedings contain papers selected for presentation in technical sessions of ICON-2020. We are thankful to our excellent team of reviewers from all over the globe who deserve full credit for the hard work of reviewing the high-quality submissions with rich technical content. From 130 submissions, 66 papers were selected, 29 long papers, 34 short papers, 3 doctoral consortium papers, representing a variety of new and interesting developments, covering a wide spectrum of NLP areas and core linguistics. Besides presentations, the conference also hosted 2 tutorials, 1 workshop, 3 shared tasks, and 18 system demonstrations.

We are deeply grateful to Prof. David Yarowsky from John Hopkins University (USA), Prof. Iryna Gurevych from Technische Universität Darmstadt (Germany), and Prof. Eduard Hovy from Carnegie Mellon University for giving the keynote lectures at ICON. We also extend our heartfelt thanks to Dr Soujanya Poria, Singapore University of Technology and Design, Singapore for giving the invited talk at

ICON.

We thank all the area chairs for the various tracks at ICON 2020, especially, Sobha Lalitha Devi (Language Resources, NLP Language Documentation and Preservation), Ashwini Vaidya, Pawan Goyal (Syntax and Lexical Semantics), Praveen Kumar G S (Named Entity Recognition, Question Answering, Information Extraction, Dialogue Systems), Amitava Das, Radhika Mamidi (Sentiment and Emotion Analysis), Karunesh Arora, Sandipan Dandapat (Machine Translation), Vasudeva Varma, Dipankar Das (Summarization, Natural Language Generation, Information Retrieval and Text Mining), C V Jawahar (Multimodality, Speech Recognition, Speech Synthesis), Raksha Sharma, Nikesh Garera, (NLP for Digital Humanities, NLP for Education), Samar Husain (Ethics in NLP, Cognitive Modelling and Psycholinguistics), Karthik Sankaranarayanan, Ashutosh Modi (Machine Learning Applications to NLP, Interpretability and Explainability of NLP models). We also thank Gurpreet Singh Lehal, Sanjay Dwivedi, Rajeev R R, Sanjeev Gupta, Neeraj Mogla, Amba Kulkarni (Co-Chairs, Tools Contest), Sudip Kumar Naskar, Sriparna Saha (Co-Chairs, Workshop/Tutorial), Preethi Jyothi (Doctoral Consortium Chair) for taking the responsibilities of the events.

We are thankful to the team members of the Artificial Intelligence-Natural Language Processing-Machine Learning (AI-NLP-ML) Group of the Department of Computer Science and Engineering for making the organization of the event at the Indian Institute of Technology Patna (IIT Patna) a success. We heartily express our gratitude to Pushpak Bhattacharyya, Asif Ekbal, Sriparna Saha, Soumitra Ghosh, Ratnesh Joshi, Prabhat Kumar Bharti, Gitanjali Singh, Tirthankar Ghosal, Apoorva Singh, and other AI-NLP-ML team members at IIT Patna for their timely help with sincere dedication to make this conference a success. We also thank and all those who came forward to help us with this task.

Finally, we thank all the researchers who responded to our call for papers and all the participants of ICON-2020, without whose overwhelming response the conference would not have been a success. We wholeheartedly thank all the reviewers who accepted our invitation and spent their valuable time reviewing the papers to maintain their high international standards. We thank the session chairs for finding out time for our conference.

December 2020
Patna

Pushpak Bhattacharyya-PC Co-chair
Dipti Misra Sharma-PC Co-chair
Rajeev Sangal-General Chair

Asif Ekbal-Organizing Committee Chair

Conference General Chair

Rajeev Sangal, IIIT Hyderabad, India

Program Chairs:

Dipti Misra Sharma, IIIT Hyderabad, India (Co-Chair)

Pushpak Bhattacharyya, IIT Bombay, India (Co-Chair)

Organizing Chair:

Asif Ekbal, IIT Patna, India

Program Committee:

Sobha Lalitha Devi, AU-KBC Research Centre, Anna University

Ashwini Vaidya, IIT Delhi

Pawan Goyal, IIT Kharagpur, India

Praveen Kumar G S, Samsung

Amitava Das, Wipro AI Lab

Radhika Mamidi, IIIT Hyderabad, India

Karunesh Arora, CDAC

Sandipan Dandapat, Microsoft

Vasudeva Varma, IIIT Hyderabad

Pushpak Bhattacharyya, IIT Bombay, India

Dipti Misra Sharma, IIIT Hyderabad

Dipankar Das, Jadavpur University

C V Jawahar, IIIT Hyderabad

Raksha Sharma, IIT Roorkee

Samar Husain, Indian Institute of Technology Delhi, India

Nikesh Garera, Flipkart

Karthik Sankaranarayanan, IBM Reseach

Ashutosh Modi, Indian Institute of Technology Kanpur

Tanmoy Charaborty, IIIT Delhi

Sriparna Saha, IIT Patna, India

Anil Kumar Vuppala, IIIT Hyderabad

Aditya Joshi, CSIRO

Girish Palshikar, Tata Consultancy Services Limited

Manish Srivastava, IIIT Hyderabad

Vishal Goyal, Punjabi University, Patiala

Sudip Kumar Naskar, Jadavpur University

Sudeshna Sarkar, IIT Kharagpur, India

Anoop Kunchukuttan, Microsoft AI and Research

Shad Akhtar, IIIT Delhi

Tools Contest Chairs:

Gurpreet Singh Lehal, Punjabi University Patiala
Sanjay Dwivedi, Central University, Lucknow
Rajeev R R, ICFOSS, Trivandrum
Sanjeev Gupta, Google, Bangalore
Neeraj Mogla, Flipkart, USA
Amba Kulkarni, University of Hyderabad

Workshop/Tutorial Chairs:

Sudip Kumar Naskar, Jadavpur University
Sriparna Saha, IIT Patna, India

Doctoral Consortium Chairs:

Preethi Jyothi, IIT Bombay, India

Invited Speakers:

Prof. David Yarowsky, John Hopkins University, USA
Prof. Iryna Gurevych, Technische Universität Darmstadt, Germany
Prof. Eduard Hovy, Carnegie Mellon University
Dr. Soujanya Poria, Singapore University of Technology and Design, Singapore

Referees

We gratefully acknowledge the excellent quality of refereeing we received from the reviewers. We thank them all for being precise and fair in their assessment and for reviewing the papers in time.

Muhammad Abulaish, South Asian University
Sriparna Saha, IIT Patna, India
Zishan Ahmad, IIT Patna, India
Md. Shad Akhtar, Indraprastha Institute of Information Technology, Delhi
Ashish Anand, Indian Institute of Technology Guwahati, India
C Anantaram, Indraprastha Institute of Information Technology, Delhi
Mohd Zeeshan Ansari, Jamia Millia Islamia
Karunesh Arora, CDAC
Rakesh Balabantaray, IIIT Bhubaneswar
Kalika Bali, Microsoft Research Labs
Somnath Banerjee, University of Milano-Bicocca
Srinivas Bangalore, Interactions Corp
Debajyoty Banik, IIT Patna, India
Biswan Barik, GE Global Research
Kingshuk Basak, Samsung Research and Development
Brendan Bena, Drury University
Riyaz A. Bhat, Interactions LLC
Meher Bhatia, IIIT-Delhi, India
Michael Carl, Kent State University
Tanmoy Chakraborty, Indraprastha Institute of Information Technology Delhi (IIIT-D), India
Sanjay Chatterji, Indian Institute Of Information Technology Kalyani
Dushyant Singh Chauhan, IIT Patna, India
Kushal Chawla, University of Southern California
Soumya Chennabasavaraj, Flipkart
Manoj Chinnakotla, Microsoft
Monojit Choudhury, Microsoft Research
Thomas Conley, University of Colorado, Colorado Springs
Sandipan Dandapat, Microsoft
Dipankar Das, Jadavpur University
Amitava Das, Wipro AI Lab
Kumar Gourav Das, Jadavpur University
Ayan Das, IIT Kharagpur, India
Niladri Sekhar Dash, Linguistic Research Unit, Indian Statistical Institute, Kolkata
Arkadia De, Indian Institute of Technology Hyderabad, India
Alok Debnath, International Institute of Information Technology, Hyderabad, India
Kuntal Dey, Accenture Technology Labs
Gihan Dias, University of Moratuwa
Sri Harsha Dumpala, Vector Institute, and Dalhousie University
Pratik Dutta, IIT Patna, India
Hridoy Sankar Dutta, IIIT Delhi, India
Indranil Dutta, Jadavpur University
Mohd Fazil, Madanapalle Institute of Technology & Science, Madanapalle, Chittoor, India
Mauajama Firdaus, IIT Patna, India

Suryakanth V Gangashetty, KLEF Deemed to be University, Green Fields
Nikesh Garera, Flipkart
Ibrahim Gashaw, Mangalore University
Tirthankar Ghosal, IIT Patna, India
Deepanway Ghosal, Singapore University of Technology and Design
Soumitra Ghosh, IIT Patna, India
Prabhat Kumar Bharti, IIT Patna, India
Sanjukta Ghosh, IIT BHU, India
Souvick Ghosh, San Jose State University
Pranav Goel, Wadhvani Institute for Artificial Intelligence
Vishal Goyal, Punjabi University Patiala
Pawan Goyal, IIT Kharagpur, India
Praveen Kumar GS, Samsung
Kamal Kumar Gupta, IIT Patna, India
Deepak Gupta, IIT Patna, India
Harald Hammarstrom, Uppsala University
Rejwanul Haque, Dublin City University
Mohammed Hasanuzzaman, Cork Institute of Technology, Dublin, Ireland
Samar Husain, Indian Institute of Technology Delhi, India
Nikhil Jaiswal, TCS Research
C V Jawahar, IIIT Hyderabad
Girish Jha, Jawaharlal Nehru University
Saurav Jha, University of Lorraine
Harimohan Jha, IIT Kharagpur
Aditya Joshi, CSIRO
Nilesh Joshi, IIT Bombay, India
Preethi Jyothi, IIT Bombay, India
Mitesh M. Khapra, IIT Madras, India
Sunil Kumar Kopparapu, TCS Research, and Innovation, Mumbai
Alapan Kuila, IIT Kharagpur, India
Amba Kulkarni, University of Hyderabad,
Malhar Kulkarni, IIT Bombay, India
Pranaw Kumar, CDAC Mumbai
Niraj Kumar, Senior Researcher, and Manager, Samsung Research Institute India, Bangalore
Ritesh Kumar Dept. of Linguistics, Dr. Bhimrao Ambedkar University, Agra
Abhinav Kumar, NIT Patna, India
Abhishek Kumar, IIT Patna, India
Vijay Kumar, Ministry of Electronics and Information Technology
Anil Kumar Singh, IIT BHU Varanasi, India
Divya Kumari, IIT Patna, India
Rina Kumari, IIT Patna, India
Anoop Kunchukuttan, Microsoft AI and Research
Bibekananda Kundu, Centre for Development of Advanced Computing (CDAC) Kolkata
Sobha Lalitha Devi, AU-KBC Research Centre, Anna University
Gurpreet Lehal, Punjabi University
Abhijith Madan, International Institute of Information Technology, Bangalore, India
Avinash Madasu, Samsung R & D Institute Bangalore
Sainik Mahata, Jadavpur University

Abhra Majumdar, IIT Kharagpur, India
Shrikant Malviya, IIIT Allahabad, India
Radhika Mamidi, IIIT Hyderabad, India
Soumik Mandal, Rutgers University
Soumil Mandal, SRM University
Pruthwik Mishra IIIT, Hyderabad
Sayantan Mitra, IIT Patna, India
Dr. Vinay Kumar Mittal, Professor, KL University
Ashutosh Modi, Indian Institute of Technology Kanpur
Aditya Mogadala, Saarland University
Vandan Mujadia, IIIT-Hyderabad
Animesh Mukherjee, IIT Kharagpur
Aditi Mukherjee, IIIT-Hyderabad
Siddhartha Mukherjee, Samsung R&D Institute India, Bangalore
Abhijith Athreya Mysore Gopinath, Pennsylvania State University
Abhishek Narayanan, Department of Computer Science and Engineering, PES University
Mukuntha Narayanan SundararamanNarayanan Sundararaman, IIT Patna, India
Sudip Kumar Naskar, Jadavpur University
Tapas Nayak, National University of Singapore
Hamada Nayel, Benha University
Anish Nediyanath, Samsung R&D Institute India - Bangalore
Vasudevan Nedumpozhimana, TU Dublin
Preksha Nema, IIT Madras
Kishorjit Nongmeikapam, Indian Institute of Information Technology(IIIT) Manipur
Deepak P, Queen's University Belfast
Jisha P Jayan, IIITMK
Partha Pakray, National Institute of Technology Silchar
Santanu Pal, Saarland University
Girish Palshikar, Tata Consultancy Services Limited
Rrubaa Panchendrarajan, National University of Singapore
Shantipriya Parida, Idiap Research Institute
Md. Aslam Parwez, Jamia Millia Islamia
Tanvina Patel, Cogknit Semantics
Kevin Patel, IIT Bombay
Sangameshwar Patil, TRDDC, TCS Research, and Innovation
Braja Gopal Patra, Department of Population Health Sciences, Weill Cornell Medicine
Abhipsa Patro, IIIT Bhubaneswar
Sayanta Paul, Indian Institute of Management
Jyoti Pawar, Goa University, Goa
Sachin Pawar, Tata Consultancy Services Ltd.
Jerin Philip, Naver Labs Europe
Soujanya Poria, Singapore University of Technology and Design
Suhan Prabhu, International Institute of Information Technology, Hyderabad
Ganesh Prasad, Samsung
S R Mahadeva Prasanna, IIT Dharwad
Michal Ptaszynski, Kitami Institute of Technology
Srinivas PYKL, IIIT Sri City
Vartika Rai, IIIT Hyderabad

Sai Krishna Rallabandi, Carnegie Mellon University
Surangika Ranathunga, university of moratuwa
Hanumant Redkar, IIT Bombay
Pattabhi RK Rao, AU-KBC Research center
Paolo Rosso, Universitat Politecnica de Valencia
Vijay Rowtula, International Institute of Information Technology, Hyderabad
Pradeep Kumar Roy, Vellore Institute of Technology, Vellore
Aniruddha Roy, IIT Kharagpur
Sowmya S Sundaram, IIT Madras
Tulika Saha, IIT Patna
Atanu Saha, Jadavpur university
Sujan Kumar Saha, Dept. of CSE, Birla Institute of Technology Mesra
Saumajit Saha, TCS Research and Innovation Labs
Sovan Kumar Sahoo, IIT Patna
Pracheta Sahoo, The University of Texas at Dallas
Tanik Saikh, IIT Patna
Naveen Saini, IIT Patna
Suyash Sangwan, IIT Patna
Karthik Sankaranarayanan, IBM Reseach
Sebastin Santy, Microsoft Research
Kamal Sarkar, Computer Science and Engineering Department, Jadavpur University
Sandip Sarkar, Hijli College
Sunil Saumya, IIIT Dharwad
Moritz Schaeffer, Johannes Gutenberg University of Mainz
Peter Scharf, International Institute of Information Technology; Indian Institute of Advanced Study
Sanket Shah, MSRI
Raksha Sharma, IIT Roorkee
Ravi Shekhar, Queen Mary University of London
Manish Shrivastava, International Institute of Information Technology Hyderabad
Smriti Singh, Samsung Research UK
Thoudam Doren Singh, NIT Silchar
Sandhya Singh, IITB
Vikram Singh, IIT Patna
Jyoti Prakash Singh, NIT Patna
Pardeep Singh, Jawaharlal Nehru University
Manjari Sinha, IITKharagpur
Sunayana Sitaram, Microsoft Research India
Saurabh Srivastava, TCS Research
Keh-Yih Su, Institute of Information Science, Academia Sinica
Chanchal Suman, IIT Patna
Bapi Raju Surampudi, International Institute of Information Technology Hyderabad
Partha Talukdar, Indian Institute of Science
Gaurish Thakkar, University of Zagreb
Anil Thakur, IIT (BHU) Varanasi
Nidhi Thakur, IIT Patna
Medari Tham, Assam Don Bosco University
Uthayasanker Thayasivam, University of Moratuwa
Uma Shanker Tiwary, IIIT Allahabad
Prajna Upadhyay, IIT Delhi

Ashwini Vaidya, IIT Delhi
Shalaka Vaidya, IIIT Hyderabad
Vasudeva Varma, IIIT Hyderabad
Deeksha Varshney, IIT Patna
Ashraf Kamal, Jamia Millia Islamia
Sriram Venkatapathy, Amazon
Samudra Vijaya, IIT Guwahati
Anil Kumar Vuppala, IIIT Hyderabad
Saumitra Yadav, International Institute of Information Technology, Hyderabad

Organized by:



Indian Institute of Technology Patna



**Natural Language Processing
Association of India**

Natural Language Processing Association of India

Table of Contents

<i>The WEAVE Corpus: Annotating Synthetic Chemical Procedures in Patents with Chemical Named Entities</i>	
Ravindra Nittala and Manish Shrivastava	1
<i>Increasing accuracy of a semantic word labelling tool based on a small lexicon</i>	
Hugo Sanjurjo-González	10
<i>Treatment of optional forms in Mathematical modelling of Pāṇini</i>	
Anupriya Aggarwal, Malhar Kulkarni	15
<i>Automatic Hadith Segmentation using PPM Compression</i>	
Taghreed Tarmom, Eric Atwell and Mohammad Alsalka	22
<i>Using multiple ASR hypotheses to boost i18n NLU performance</i>	
Charith Peris, Gokmen Oz, Khadige Abboud, Venkata sai Varada, Prashan Wanigasekara and Haidar Khan	30
<i>A Grammatical Sketch of Asur: A North Munda language</i>	
Zoya Khalid	40
<i>English to Manipuri and Mizo Post-Editing Effort and its Impact on Low Resource Machine Translation</i>	
Loitongbam Sanayai Meetei, Thoudam Doren Singh, Sivaji Bandyopadhyay, Mihaela Vela and Josef van Genabith	50
<i>Learning to Interact: An Adaptive Interaction Framework for Knowledge Graph Embeddings</i>	
Chandrasah ., Nilesh Agrawal and Partha Talukdar	60
<i>Inducing Interpretability in Knowledge Graph Embeddings</i>	
Chandrasah ., Tathagata Sengupta, Cibi Pragadeesh and Partha Talukdar	70
<i>Solving Arithmetic Word Problems Using Transformer and Pre-processing of Problem Texts</i>	
Kaden Griffith and Jugal Kalita	76
<i>Clickbait in Hindi News Media : A Preliminary Study</i>	
Vivek Kaushal and Kavita Vemuri	85
<i>Self Attended Stack-Pointer Networks for Learning Long Term Dependencies</i>	
Salih Tuç and Burcu Can	90
<i>Creation of Corpus and Analysis in Code-Mixed Kannada-English Social Media Data for POS Tagging</i>	
Abhinav Reddy Appidi, Vamshi Krishna Srirangam, Darsi Suhas and Manish Shrivastava	101
<i>Identifying Complaints from Product Reviews in Low-resource Scenarios via Neural Machine Translation</i>	
Raghvendra Pratap Singh, Rejwanul Haque, Mohammed Hasanuzzaman and Andy Way	108
<i>Generative Adversarial Networks for Annotated Data Augmentation in Data Sparse NLU</i>	
Olga Golovneva and Charith Peris	117
<i>BertAA : BERT fine-tuning for Authorship Attribution</i>	
Maël Fabien, Esaú Villatoro-Tello, Petr Motlicek and Shantipriya Parida	127

<i>TREE ADJOINING GRAMMAR BASED "LANGUAGE INDEPENDENT GENERATOR"</i>	
Pavan Kurariya, Prashant Chaudhary, Jahnvi Bodhankar, Lenali Singh, Ajai Kumar and Hemant Darbari	138
<i>Exploration of Cross-lingual Summarization for Kannada-English Language Pair</i>	
Vinayaka R Kamath, Rachana Aithal K R, Vennela K and Mamatha HR	144
<i>Hater-O-Genius Aggression Classification using Capsule Networks</i>	
Parth Patwa, Srinivas PYKL, Amitava Das, Prerana Mukherjee and Viswanath Pulabaigari	149
<i>A New Approach to Claim Check-Worthiness Prediction and Claim Verification</i>	
Shukrity Si, ANISHA DATTA and Sudip Kumar Naskar	155
<i>Improving Passage Re-Ranking with Word N-Gram Aware Coattention Encoder</i>	
Chaitanya Alaparathi and Manish Shrivastava	161
<i>Language Model Metrics and Procrustes Analysis for Improved Vector Transformation of NLP Embeddings</i>	
Thomas Conley and Jugal Kalita	170
<i>Cognitively Aided Zero-Shot Automatic Essay Grading</i>	
Sandeep Mathias, Rudra Murthy, Diptesh Kanojia and Pushpak Bhattacharyya	175
<i>Automated Arabic Essay Evaluation</i>	
Abeer Alqahtani and Amal Alsaif	181
<i>Semantic Extractor-Paraphraser based Abstractive Summarization</i>	
Anubhav Jangra, Raghav Jain, Vaibhav Mavi, Sriparna Saha and Pushpak Bhattacharyya	191
<i>ThamizhiUDp: A Dependency Parser for Tamil</i>	
Kengatharaiyer Sarveswaran and Gihan Dias	200
<i>Constructing a Korean Named Entity Recognition Dataset for the Financial Domain using Active Learning</i>	
Dong-Ho Jeong, Min-Kang Heo, Hyung-Chul Kim and Sang-Won Park	208
<i>Self-Supervised Claim Identification for Automated Fact Checking</i>	
Archita Pathak, Mohammad Abuzar Shaikh and Rohini Srihari	213
<i>SUKHAN: Corpus of Hindi Shayaris annotated with Sentiment Polarity Information</i>	
Salil Aggarwal, Abhigyan Ghosh and Radhika Mamidi	228
<i>Improving Neural Machine Translation for Sanskrit-English</i>	
Ravneet Punia, Aditya Sharma, Sarthak Pruthi and Minni Jain	234
<i>Parsing Indian English News Headlines</i>	
Samapika Roy, Sukhada and Anil Kumar Singh	239
<i>WORD SENSE DISAMBIGUATION FOR KASHMIRI LANGUAGE USING SUPERVISED MACHINE LEARNING</i>	
Tawseef Ahmad Mir and Aadil Ahmad Lawaye	243
<i>Sentimental Poetry Generation</i>	
Kasper Aalberg Rostvøld and Björn Gambäck	246

<i>WEKA in Forensic Authorship Analysis: A corpus-based approach of Saudi Authors</i> Mashaël M. AlAmr and Eric Atwell	257
<i>Native-Language Identification with Attention</i> Stian Steinbakken and Björn Gambäck	261
<i>Does a Hybrid Neural Network-based Feature Selection Model Improve Text Classification?</i> Suman Dowlagar and Radhika Mamidi	272
<i>Efforts Towards Developing a Tamang Nepali Machine Translation System</i> Binaya Kumar Chaudhary, Bal Krishna Bal and Rasil Baidar	281
<i>Event Argument Extraction using Causal Knowledge Structures</i> Debanjana Kar, Sudeshna Sarkar and Pawan Goyal	287
<i>Claim extraction from text using transfer learning.</i> Acharya Ashish Prabhakar, Salar Mohtaj and Sebastian Möller	297
<i>Assamese Word Sense Disambiguation using Genetic Algorithm</i> Arjun Gogoi, Nomi Baruah and Shikhar Kr. Sarma	303
<i>Free Word Order in Sanskrit and Well-nestedness</i> Sanal Vikram and Amba Kulkarni	308
<i>A Multi-modal Personality Prediction System</i> Chanchal Suman, Aditya Gupta, Sriparna Saha and Pushpak Bhattacharyya	317
<i>D-Coref: A Fast and Lightweight Coreference Resolution Model using DistilBERT</i> Chanchal Suman, Jeetu Kumar, Sriparna Saha and Pushpak Bhattacharyya	323
<i>Semantic Slot Prediction on low corpus data using finite user-defined list</i> Bharatram Natarajan, Dharani Simma, Chirag Singh, Anish Nediyanath and Sreoshi Sengupta	329
<i>Leveraging Latent Representations of Speech for Indian Language Identification</i> Samarjit Karmakar and P Radha Krishna	334
<i>Acoustic Analysis of Native (L1) Bengali Speakers' Phonological Realization of English Lexical Stress Contrast</i> Shambhu Nath Saha and Shyamal Kr. Das Mandal	341
<i>Towards Performance Improvement in Indian Sign Language Recognition</i> Kinjal Mistree, Devendra Thakor and Brijesh Bhatt	349
<i>Question and Answer pair generation for Telugu short stories</i> Meghana Bommadi, Shreya Terupally and Radhika Mamidi	355
<i>Detection of Similar Languages and Dialects Using Deep Supervised Autoencoder</i> Shantipriya Parida, Esaú Villatoro-Tello, Sajit Kumar, Maël Fabien and Petr Motlicek	362
<i>Weak Supervision using Linguistic Knowledge for Information Extraction</i> Sachin Pawar, Girish Palshikar, Ankita Jain, Jyoti Bhat and Simi Johnson	368
<i>Leveraging Alignment and Phonology for low-resource Indic to English Neural Machine Transliteration</i> Parth Patel, Manthan Mehta, Pushpak Bhattacharyya and Arjun Atreya	373

<i>STHAL: Location-mention Identification in Tweets of Indian-context</i> Kartik Verma, Shobhit Sinha, Md. Shad Akhtar and Vikram Goyal	379
<i>On-Device detection of sentence completion for voice assistants with low-memory footprint</i> Rahul Kumar, Vijeta Gour, Chandan Pandey, Godawari Sudhakar Rao, Priyadarshini Pai, Anmol Bhasin and Ranjan Samal	384
<i>Polarization and its Life on Social Media: A Case Study on Sabarimala and Demonetisation</i> Ashutosh Ranjan, Dipti Sharma and Radhika Krishnan	393
<i>A Rule Based Lightweight Bengali Stemmer</i> Souvick Das, Rajat Pandit and Sudip Kumar Naskar	400
<i>End-to-End Automatic Speech Recognition for Gujarati</i> Deepang Raval, Vyom Pathak, Muktan Patel and Brijesh Bhatt	409
<i>Deep Neural Model for Manipuri Multiword Named Entity Recognition with Unsupervised Cluster Feature</i> Jimmy Laishram, Kishorjit Nongmeikapam and Sudip Kumar Naskar	420
<i>ScAA: A Dataset for Automated Short Answer Grading of Children's free-text Answers in Hindi and Marathi</i> Dolly Agarwal, Somya Gupta and Nishant Baghel	430
<i>Exploring Pair-Wise NMT for Indian Languages</i> Karthek Akella, Sai Himall Allu, Sridhar Suresh Ragupathi, Aman Singhal, Zeeshan Khan, C.V. Jawahar and Vinay P. Namboodiri	437
<i>Only text? only image? or both? Predicting sentiment of internet memes</i> Pranati Behera, Mamta . and Asif Ekbal	444
<i>Towards Bengali Word Embedding: Corpus Creation, Intrinsic and Extrinsic Evaluations</i> Md. Rajib Hossain and Mohammed Moshuiul Hoque	453
<i>Annotated Corpus of Tweets in English from Various Domains for Emotion Detection</i> Soumitra Ghosh, Asif Ekbal, Pushpak Bhattacharyya, Sriparna Saha, Vipin Tyagi, Alka Kumar, Shikha Srivastava and Nitish Kumar	460
<i>PhraseOut: A Code Mixed Data Augmentation Method for Multilingual Neural Machine Translation</i> Binu Jasim, Vinay Namboodiri and C V Jawahar	470
<i>CLPLM: Character Level Pretrained Language Model for Extracting Support Phrases for Sentiment Labels</i> Raj Pranesh, Sumit Kumar and Ambesh Shekhar	475
<i>Developing a Faroese PoS-tagging solution using Icelandic methods</i> Hinrik Hafsteinsson and Anton Karl Ingason	481
<i>Leveraging Multi-domain, Heterogeneous Data using Deep Multitask Learning for Hate Speech Detection</i> Prashant Kapil and Asif Ekbal	491

Conference Program

Day 1: Saturday, December 19, 2020

+ 10:00 - 11:00 **Inaugural Ceremony**

+ 11:30 - 13:00 BREAK

+ 13:00 -14:30 **Technical Session I: Information Extraction-I**

Session Chair: Karthik Sankaranarayanan

Automatic Hadith Segmentation using PPM Compression

Taghreed Tarmom, Eric Atwell and Mohammad Alsalka

Learning to Interact: An Adaptive Interaction Framework for Knowledge Graph Embeddings

Chandahas, Nilesh Agrawal and Partha Talukdar

Event Argument Extraction using Causal Knowledge Structures

Debanjana Kar, Sudeshna Sarkar and Pawan Goyal

Weak Supervision using Linguistic Knowledge for Information Extraction

Sachin Pawar, Girish Palshikar, Ankita Jain, Jyoti Bhat and Simi Johnson

Technical Session II: NLP Language Documentation and Preservation

Session Chair: Sobha Lalitha Devi

A Grammatical Sketch of Asur: A North Munda language

Zoya Khalid

Treatment of optional forms in Mathematical modelling of Pāṇini

Anupriya Aggarwal and Malhar Kulkarni

Language Model Metrics and Procrustes Analysis for Improved Vector Transformation of NLP Embeddings

Thomas Conley and Jugal Kalita

Assamese Word Sense Disambiguation using Genetic Algorithm

Arjun Gogoi, Nomi Baruah and Shikhar Kr. Sarma

Technical Session III: Computational Social Science and Social Media

Session Chair: Tanmoy Chakraborty

Identifying Complaints from Product Reviews in Low-resource Scenarios via Neural Machine Translation

Raghvendra Pratap Singh, Rejwanul Haque, Mohammed Hasanuzzaman and Andy Way

Hater-O-Genius Aggression Classification using Capsule Networks

Parth Patwa, Srinivas PYKL, Amitava Das, Prerana Mukherjee and Viswanath Pulabaihari

Native-Language Identification with Attention

Stian Steinbakken and Björn Gambäck

Acoustic Analysis of Native (L1) Bengali Speakers' Phonological Realization of English Lexical Stress Contrast

Shambhu Nath Saha and Shyamal Kr. Das Mandal

+15:00-16:00 **Keynote Lecture 1:** Prof. Dr. Iryna Gurevych, Technische Universität Darmstadt, Germany

Title: Let's Argue - Understanding and Generating Natural Language Arguments

Session Chair: Sudeshna Sarkar

+16:00-16:30 **BUFFER**

Technical Session IV: Sentiment and Emotion Analysis

Session Chair: Amitava Das

Polarization and its Life on Social Media: A Case Study on Sabarimala and Demonetisation

Ashutosh Ranjan, Dipti Sharma, and Radhika Krishnan

Only text? only image? or both? Predicting sentiment of internet memes

Pranati Behera, Mamta and Asif Ekbal

Leveraging Multi-domain, Heterogeneous Data using Deep Multitask Learning for Hate Speech Detection

Prashant Kapil and Asif Ekbal

CLPLM: Character Level Pretrained Language Model for Extracting Support Phrases for Sentiment Labels

Raj Pranesh, Sumit Kumar and Ambesh Shekhar

Technical Session V: Named Entity Recognition

Session Chair: Sriparna Saha

The WEAVE Corpus: Annotating Synthetic Chemical Procedures in Patents with Chemical Named Entities

Ravindra Nittala and Manish Shrivastava

Deep Neural Model for Manipuri Multiword Named Entity Recognition with Unsupervised Cluster Feature

Jimmy Laishram, Kishorjit Nongmeikapam and Sudip Kumar Naskar

Constructing a Korean Named Entity Recognition Dataset for the Financial Domain using Active Learning

Dong-Ho Jeong, Min-Kang Heo, Hyung-Chul Kim and Sang-Won Park

Technical Session VI: Multimodality/Speech Recognition

Session Chair: Anil Kumar Vuppala

A Multi-modal Personality Prediction System

Chanchal Suman, Aditya Gupta, Sriparna Saha and Pushpak Bhattacharyya

End-to-End Automatic Speech Recognition for Gujarati

Deepang Raval, Vyom Pathak, Muktan Patel and Brijesh Bhatt

Using multiple ASR hypotheses to boost i18n NLU performance

Charith Peris, Gokmen Oz, Khadige Abboud, Venkata sai Varada, Prashan Wani-gasekara, and Haidar Khan

Leveraging Latent Representations of Speech for Indian Language Identification

Samarjit Karmakar and P Radha Krishna

+17:30-18:00 **BUFFER**

+18:00-19:00 **Keynote Lecture 2:** Prof. Eduard Hovy, Carnegie Mellon University

Title: From Simple to Complex QA

Session Chair: Prof. Pushpak Bhattacharyya

+19:00-19:30 **BUFFER**

+19:30-21:00 **NLPAI Meeting**

Day 2: Sunday, December 20, 2020

+ 11:00 -13:00 **Technical Session VII:** Information Extraction-II

Session Chair: Dipankar Das

Inducing Interpretability in Knowledge Graph Embeddings

Chandrabhas, Tathagata Sengupta, Cibi Pragadeesh and Partha Talukdar

Solving Arithmetic Word Problems Using Transformer and Pre-processing of Problem Texts

Kaden Griffith and Jugal Kalita

Generative Adversarial Networks for Annotated Data Augmentation in Data Sparse NLU

Olga Golovneva and Charith Peris

Semantic Extractor-Paraphraser based Abstractive Summarization

Anubhav Jangra, Raghav Jain, Vaibhav Mavi, Sriparna Saha and Pushpak Bhat-tacharyya

Technical Session VIII:: Machine Learning Applications to NLP-I

Session Chair: Aditya Joshi

BertAA : BERT fine-tuning for Authorship Attribution

Maël Fabien, Esaú Villatoro-Tello, Petr Motliceck and Shantipriya Parida

Claim extraction from text using transfer learning.

Acharya Ashish Prabhakar, Salar Mohtaj and Sebastian Möller

On-Device detection of sentence completion for voice assistants with low-memory footprint

Rahul Kumar, Vijeta Gour, Chandan Pandey, Godawari Sudhakar Rao, Priyadarshini Pai, Anmol Bhasin and Ranjan Samal

A New Approach to Claim Check-Worthiness Prediction and Claim Verification

Shukrity Si, ANISHA DATTA and Sudip Kumar Naskar

Technical Session IX: Machine Learning Applications to NLP-II

Session Chair: Ashutosh Modi

Clickbait in Hindi News Media : A Preliminary Study

Vivek Kaushal and Kavita Vemuri

Does a Hybrid Neural Network-based Feature Selection Model Improve Text Classification?

Suman Dowlagar and Radhika Mamidi

Semantic Slot Prediction on low corpus data using finite user-defined list

Bharatram Natarajan, Dharani Simma, Chirag Singh, Anish Nediyanath and Sreoshi Sengupta

Detection of Similar Languages and Dialects Using Deep Supervised Autoencoder

Shantipriya Parida, Esaú Villatoro-Tello, Sajit Kumar, Maël Fabien and Petr Motlicek

Sentimental Poetry Generation

Kasper Aalberg Røstvold and Björn Gambäck

Towards Performance Improvement in Indian Sign Language Recognition

Kinjal Mistree, Devendra Thakor and Brijesh Bhatt

Technical Session X: Machine Translation-I

Session Chair: Nikesh Garera

Exploring Pair-Wise NMT for Indian Languages

Kartheek Akella, Sai Himallu, Sridhar Suresh Ragupathi, Aman Singhal, Zee-shan Khan, C.V. Jawahar and Vinay P. Namboodiri

PhraseOut: A Code Mixed Data Augmentation Method for Multilingual Neural Machine Translation

Binu Jasim, Vinay Namboodiri and C V Jawahar

Efforts Towards Developing a Tamang Nepali Machine Translation System

Binaya Kumar Chaudhary, Bal Krishna Bal and Rasil Baidar

TREE ADJOINING GRAMMAR BASED "LANGUAGE INDEPENDENT GENERATOR"

Pavan Kurariya, Prashant Chaudhary, Jahnavi Bodhankar, Lenali Singh, Ajai Kumar and Hemant Darbari

+13:00-14:00 **BREAK**

+ 14:00 -16:00 **Technical Session XI: Language Resources- I**

Session Chair: Girish Palsikar

A Rule Based Lightweight Bengali Stemmer

Souvick Das, Rajat Pandit and Sudip Kumar Naskar

Towards Bengali Word Embedding: Corpus Creation, Intrinsic and Extrinsic Evaluations

Md. Rajib Hossain and Mohammed Moshikul Hoque

Annotated Corpus of Tweets in English from Various Domains for Emotion Detection

Soumitra Ghosh, Asif Ekbal, Pushpak Bhattacharyya, Sriparna Saha, Vipin Tyagi, Alka Kumar, Shikha Srivastava and Nitish Kumar

STHAL: Location-mention Identification in Tweets of Indian-context

Kartik Verma, Shobhit Sinha, Md. Shad Akhtar and Vikram Goyal

Developing a Faroese PoS-tagging solution using Icelandic methods

Hinrik Hafsteinsson and Anton Karl Ingason

Technical Session XII: Language Resources-II

Session Chair: Manish Srivastava and Vishal Goyal

Increasing accuracy of a semantic word labelling tool based on a small lexicon

Hugo Sanjurjo-González

Creation of Corpus and Analysis in Code-Mixed Kannada-English Social Media Data for POS Tagging

Abhinav Reddy Appidi, Vamshi Krishna Srirangam, Darsi Suhas and Manish Srivastava

Exploration of Cross-lingual Summarization for Kannada-English Language Pair

Vinayaka R Kamath, Rachana Aithal K R, Vennela K and Mamatha HR

SUKHAN: Corpus of Hindi Shayaris annotated with Sentiment Polarity Information

Salil Aggarwal

Technical Session XIII: NLP for Education

Session Chair: Sudip Naskar

Cognitively Aided Zero-Shot Automatic Essay Grading

Sandeep Mathias, Rudra Murthy, Diptesh Kanojia and Pushpak Bhattacharyya

Automated Arabic Essay Evaluation

Abeer Alqahtani and Amal Alsaif

Question and Answer pair generation for Telugu short stories

Meghana Bommadi, Shreya Terupally and Radhika Mamidi

ScAA: A Dataset for Automated Short Answer Grading of Children's free-text Answers in Hindi and Marathi

Dolly Agarwal, Somya Gupta and Nishant Baghel

+16:00-16:30 **BUFFER**

+16:30-17:30 **Technical Session XIV: Information Retrieval and Text Mining**

Session Chair: Sudeshna Sarkar

Improving Passage Re-Ranking with Word N-Gram Aware Coattention Encoder
Chaitanya Alaparathi and Manish Shrivastava

Self-Supervised Claim Identification for Automated Fact Checking
Archita Pathak, Mohammad Abuzar Shaikh and Rohini Srihari

D-Coref: A Fast and Lightweight Coreference Resolution Model using DistilBERT
Chanchal Suman, Jeetu Kumar, Sriparna Saha and Pushpak Bhattacharyya

Technical Session XV: Syntax
Session Chair: Pawan Goyal

Self Attended Stack-Pointer Networks for Learning Long Term Dependencies
Salih Tuc and Burcu Can

ThamizhiUDp: A Dependency Parser for Tamil
Kengatharaiyer Sarveswaran and Gihan Dias

Free Word Order in Sanskrit and Well-nestedness
Sanal Vikram and Amba Kulkarni

Technical Session XVI: Machine Translation-II
Session Chair: Anoop Kunchukuttan

English to Manipuri and Mizo Post-Editing Effort and its Impact on Low Resource Machine Translation
Loitongbam Sanayai Meetei, Thoudam Doren Singh, Sivaji Bandyopadhyay, Mihaela Vela and Josef van Genabith

Improving Neural Machine Translation for Sanskrit-English
Ravneet Punia, Aditya Sharma, Sarthak Pruthi and Minni Jain

Leveraging Alignment and Phonology for low-resource Indic to English Neural Machine Transliteration
Parth Patel and Manthan Mehta

+17:30-18:00 **BUFFER**

+18:00-19:00 **Keynote Lecture 3:** David Yarowsky, John Hopkins University, USA
Title: Translingual Learning of 1000+ Languages
Session Chair: Prof. Dipti Misra Sharma

+19:00-19:30 **Valedictory Session**

Pre-conference: Friday, December 18, 2020

+ 15:00 -16:30 Doctoral Consortium Session

Session Chair: Anil Kumar Singh

Parsing Indian English News Headlines

Samapika Roy, Sukhada and Anil Kumar Singh

WORD SENSE DISAMBIGUATION FOR KASHMIRI LANGUAGE USING SUPERVISED MACHINE LEARNING

TAWSEEF AHMAD MIR

WEKA in Forensic Authorship Analysis: A corpus-based approach of Saudi Authors

Masha'el AlAmr and Eric Atwell