ICON 2021

# 18th International Conference on Natural Language Processing

## Proceedings of the Conference

December 16 - 19, 2021
National Institute of Technology Silchar, India

©2021 NLP Association of India (NLPAI)

# 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-2021. 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 204 submissions, 78 papers were selected, 51 long papers, 27 short papers, 2 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 8 tutorials, 4 workshop, 2 shared tasks, and 3 system demonstrations.

We are deeply grateful to Prof. Josef van Genabith from DFKI and Saarland University (Germany), Prof. Philip Resnik from University of Maryland (USA), Prof. Rada Mihalcea from University of Michigan (USA) and Dr. Louis-Philippe Morency from Carnegie Mellon University (USA) for giving the keynote lectures at ICON-2021.

We thank all the area chairs for the various tracks at ICON-2021, especially, Dipti Misra Sharma, Pawan Goyal (Syntax and Semantics), Kamal Kumar Choudhary, Krishna Prasad Miyapuram (Computational Psycholinguistics), Asif Ekbal, Shubhashis Sengupta (QA, Information Extraction), Ranjani Parthasarathi, Girish K. Palshikar (Information Retrieval and Text Mining), Dipankar Das, Raksha Sharma (Sentiment Analysis and Emotion Recognition), Girish Nath Jha, Niladri Shekhar Dash (Language Resources and Evaluation), Srinivas Bangalore, Kalika Bali (Speech), C V Jawahar, Asutosh Modi (Multimodality), Anoop Kunchukuttan, Karunesh Arora (Machine Translation), Monojit Choudhury, Anand Kumar M (NLP Applications), Sudeshna Sarkar, Sandipan Dandapat (Machine Learning in NLP), Vasudeva Varma, Sriparna Saha (Natural Language Text Generation). We also thank Radhika Mamidi, Samar Husain (Doctoral Consortium), Partha Pakray, Vishal Goyal (Shared Task/Tools Contest), Sudip Kumar Naskar, Amitava Das (Workshop/Tutorial) for taking the responsibilities of the events.

We are thankful to the team members of the Centre for Natural Language Processing (CNLP) Group and the Department of Computer Science and Engineering for making the organization of the event at the National Institute of Technology Silchar (NIT Silchar) a success.

We heartily express our gratitude to Partha Pakray, Naresh Babu Muppalaneni, Badal Soni, Loitongbam Sanayai Meetei, Ringki Das, Salam Michael Singh, Alok Singh for their timely help with sincere dedication to make this conference a success. We also thank all the student volunteers 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-2021, 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 2021
Silchar

Sivaji Bandyopadhyay - PC Co-chair
Sobha L - PC Co-chair
Pushpak Bhattacharyya - General Chair

Thoudam Doren Singh - Organizing Committee Chair

**Advisory Committee Chair**

Rajeev Sangal, IIIT Hyderabad, India

**Conference General Chair**

Pushpak Bhattacharyya, IIT Bombay, India

**Program Chairs:**

Sivaji Bandyopadhyay, NIT Silchar, India (Co-Chair)
Sobha L, AU-KBC Research Centre, India (Co-Chair)

**Organizing Chair:**

Thoudam Doren Singh, NIT Silchar, India

**Program Committee:**

Dipti Misra Sharma, IIIT Hyderabad
Pawan Goyal, IIT Kharagpur
Kamal Kumar Choudhary, IIT Ropar
Krishna Prasad Miyapuram, IIT Gandhinagar
Asif Ekbal, IIT Patna
Shubhashis Sengupta, Accenture
Ranjani Parthasarathi, Anna University
Girish K. Palshikar, TCS Innovation
Dipankar Das, Jadavpur University
Raksha Sharma, IIT Roorkee
Girish Nath Jha, JNU
Niladri Shekhar Dash, ISI Kolkata
Srinivas Bangalore, Interactions LLC, USA
Kalika Bali, Microsoft Research India
C V Jawahar, IIIT Hyderabad
Asutosh Modi, IIT Kanpur
Anoop Kunchukuttan, Microsoft
Karunesh Arora, CDAC Noida
Monojit Choudhury, Microsoft
Anand Kumar M, NIT Surathkal
Sudeshna Sarkar, IIT Kharagpur
Sandipan Dandapat, Microsoft
Vasudeva Varma, IIIT Hyderabad
Sriparna Saha, IIT Patna

**Doctoral Consortium Chairs**

Radhika Mamidi, IIIT Hyderabad
Samar Husain, IIT Delhi

**Shared Task/Tools Contest Chairs**

Partha Pakray, NIT Silchar
Vishal Goyal, Punjabi University, India

**Workshop/Tutorial Chairs**

Sudip Kumar Naskar, Jadavpur University
Amitava Das, Wipro AI

**Invited Speakers**

Prof. Josef van Genabith, DFKI and Saarland University, Germany
Prof. Philip Resnik, University of Maryland, USA
Prof. Rada Mihalcea, University of Michigan, USA
Dr. Louis-Philippe Morency, Carnegie Mellon University,USA

# 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.

Abhay Garg, TCS Research Pune, India
Abhijith Athreya Mysore Gopinath, Pennsylvania State University, USA
Abir Chakraborty, Microsoft India
Abu Kaisar Mohammad Masum, Daffodil International University
Aditya Joshi, CSIRO, India
Aditya Mogadala, Saarland University, Germany
Aditya Shah, IIT Indore, India
Aditya Srivastava, IIIT Hydebad, India
Adwait Ratnaparkhi, University of California Santa Cruz, USA
Adway Mitra, IIT Kharagpur
Aiusha Hujon, NEHU, Shillong
Akshay Parekh, IITGuwahati
Àlex Atrio, HEIG-VD and EPFL, Switzerland
Ali Saeedi, Kent State University
Alok Singh, NIT Silchar
Amba Kulkarni, University of Hyderabad, India
Amit Kumar, Université de Caen
Amitava Das, Wipro AI Lab, Bangalore, India
Amrit Verma, Siemens, India
Anand Kumar M, NIT Surathkal
Andrei Popescu-Belis, HEIG-VD and EPFL, Switzerland
Angel De Paula Magnossão, PRHLT, Spain
Anmol Nayak, ARiSE Labs, Bosch
Anupam Jamatia, NIT Agartala
Anupam Mondal, Jadavpur University, India
Anurag Dwarakanath, Amazon, Bangalore, India
Anustup Bhattacharya,
Apoorva Singh, IIT Patna
Arnab Bhattacharya, IIT Kanpur
Ashis Chanda, Temple University, USA
Ashraf Kamal, ACL Digital, India
Asif Ekbal, IIT Patna
Atul Kumar, Roku, USA
Ayan Das, BITS Pilani, Hyderbad
Badal Soni, NIT Silchar
Bharatram Natarajan, SRIB, India
Bibekananda Kundu, CDAC,Kolkata
Bishwa Das, Maharaja Sriram Chandra Bhanja Deo University
Björn Gambäck, Norwegian University of Science and Technology; RISE, Sweden
Bo Dong, University of Texas at Dallas
Braja Gopal Patra, Weill Cornell Medicine, New York, USA
Bratati Dhar, ISI Kolkata
Brijesh Bhatt, Dharmsinh Desai University, Nadiad
Carlos Escolano, Universitat Politècnica de Catalunya, Barcelona
Chandni Saxena, The Chinese University of Hong Kong, Shatin, NT, Hong Kong
Chandrahas, IIS Bengaluru, India
Chandresh Maurya, IIT Indore

Chayanika Chatterjee,
Chenchen Ding, National Institute of Information and Communications Technology, Japan
Chiranjeevi Yarra, IIIT Hyderabad
Cristina España-Bonet, DFKI and Saarland University
Danish Ebadulla, PES University Bangalore, India
Deeksha Varshney, IIT Patna
Deepali Jain, NIT Silchar
Dipankar Das, Jadavpur University, Kolkata
Diptesh Kanojia, University of Surrey, United Kingdom
Ganesh Mirishkar, IIIT,Hyderabad
Gatha Sharma, Shiv Nadar University
Gihan Dias, University of Moratuwa, Sri Lanka
Girish Jha, Jawaharlal Nehru University, India
Gouri Sankar Majumder, IIT Guwahati
Gunjan Ansari, JSS ATE-Noida
Hanumant Redkar, IIT Bombay, India
Himangshu Sarma, IIIT Sri City, Chittoor
Hour Kaing, National Institute of Information and Communications Technology, Japan
Hridoy Sankar Dutta, University of Cambridge
Indranil Roy, Jadavpur University
Jingxuan Yang, Beijing University
Joel Beckles, University of the West Indies
Josef van Genabith, DFKI and Saarland University
Juan Javier Sánchez Junquera, Universitat Politècnica de València
Jyotisankar Kalia, Gandhi Institute for Education and Technology (GIET), Khurda, Bhubaneswar
K. Sreenivasa Rao, IIT Kharagpur
Kamal Sarkar, Jadavpur University
Kanika Kalra, TCS Research Pune
Katsuhito Sudoh, Nara Institute of Science and Technology, Japan
Kaustubh Agarwal, Netaji Subhas University of Technology New Delhi
Kengatharaiyer Sarveswaran, University of Jaffna, Sri Lanka
Komal Bharti, IIT Guwahati
Loitongbam Sanayai Meetei, NIT Silchar
Lov Kumar, BITS Pilani Hyderabad
Lukas Edman, University of Groningen
Malhar Kulkarni, IIT Bombay
Manish Verma, IBM India
Mauajama Firdaus, IIT Patna
Mayur Patidar, TCS Research Pune
Meghna Majumder
Michael Carl, Kent State University, USA
Minghan Wang, Huawei
Minh Thuan Nguyen, University of Engineering and Technology, VNU Hanoi
Mithun Kumar S R, BITS Pilani
Mohd Fazil, MITS Madanapalle
Mohd Zeeshan Ansari, Jamila Mia Islamia
Monalisa Dey, Jadavpur University
Mukesh Kumar Rohil, BITS Pilani
Musica Supriya, Manipal Institute of Technology, India
Muskan Garg, Thapar Institute of Engineering and Technology, India
Ngo Xuan Bach, Posts and Telecommunications Institute of Technology (PTIT), Hanoi, Vietnam

Nidhi Arora, Interactions
Niladri Dash, ISI Kolkata
Niraj Kumar, Samsung
Nirnoy Dutta
Nitin Ramrakhiyani, Tata Research Development and Design Centre (TRDDC) in Pune
Niyati Bafna, Faculty of Mathematics and Physics, Charles University
Paolo Rosso, Universitat Politècnica de València
Parameswari Krishnamurthy, University of Hyderabad
Partha Pakray, NIT Silchar
Peter Scharf, Indian Institute of Advanced Study, Shimla
Pracheta Sahoo, University of Texas at Dallas
Pradeep Kumar Roy, IIIT Surat
Prajit Dhar, University of Groningen
Prajna Upadhyay, Research Engineer, CEDAR Team, INRIA, France
Pramit Bhattacharyya, IIT Kanpur
Pranav Goel, Wadhwani AI
Pritam Deka, Queen's University Belfast, UK
Priyatam Naravajhula, CBIT, Hyderabad
Punnoose A K, Flare Speech Systems
Rabul Laskar, NIT Silchar, India
Radha Krishna Guntur, VNRVJIET, Hyderabad
Rajaswa Patil, TCS Research - Tata Consultancy Services Limited
Rajat Kumar, TCS Research
Rajendra Roul, Thapar Institute of Engineering and Technology
Raksha Sharma, IIT Roorkee
Ramakrishna Appicharla, Indian Institute of Technology Patna
Ramesh Dadi, SR University, India
Ramya Akula, University of Central Florida
Ranjani Parthasarathi, Anna University
Ratnakar Pawar, Sony Playstation ML / Illinois Tech Chicago
Ravi Shekhar, Queen Mary University of London
Redwan Rizvee, Department of Computer Science and Engineering, University of Dhaka
Reynier Ortega Bueno, PRHLT Research Center, UPV
Riccardo Cervero, Università degli Studi di Milano Bicocca, Italy
Rina Kumari, IIT Patna
Ringki Das, NIT Silchar
Ritesh Kumar, Dr. Bhimrao Ambedkar University
Rusa Bhowmik, Jadavpur University
Sachin Pawar, TCS Research, Pune
Sadat Shahriar, University of Houston
Sai Muralidhar Jayanthi, MindMeld, Cisco Systems Inc.
Sainik Mahata, Institute of Engineering and Management
Salam Michael Singh, NIT Silchar
Samah Alazani, Dr. Babasaheb Ambedkar Marathawada University, Aurangabad, India
Samir Karmakar, Jadavpur University
Samudra Vijaya, IIT Guwahati
Sandhya Singh, CFILT, IIT Bombay
Sandip Sarkar, Jadavpur University
Sangameshwar Patil, TCS and Doctoral Research Scholar, IIT Madras
Sangram Kapre, Playstation
Sanjukta Ghosh, IIT BHU

Santanu Pal, Wipro AI
Santosh Kumar Mishra, Indian Institute of Technology Patna
Sara Renjit, Cochin University of Science and Technology, India
Satoshi Nakamura, Nara Institute of Science and Technology, Japan
Saujas Vaduguru, IIIT Hyderabad
Saumajit Saha, TCS Research
Seba Susan, Delhi Technological University
Shaily Bhatt, Predoctoral Researcher, Google Research
Shantipriya Parida, Silo AI
Shubhashis Sengupta, Accenture Labs
Siddhartha Mukherjee, SRIB India
Sixing Lu, Amazon Alexa AI
Soma Paul, IIIT-Hyderabad
Soumik Mandal, Ohio State University
Soumitra Ghosh, IIT Patna
Soumyadeep Ghosh, Mastercard
Sovan Kumar Sahoo, Indian Institute of Technology Patna
Srija Deb, ISI Kolkata
Sripathi Sripada, School of Vedic Sciences, MIT-ADT University
Sudeshna Das, IIT Kharagpur
Sukhada, IIT (BHU), Varanasi
Sunil Kumar Kopparapu, Tata Consultancy Services Ltd., India.
Surangika Ranathunga, University of Moratuwa
Surmila Thokchom, NIT Meghalaya
Svetla Koeva, IBL, Bulgarian Academy of Sciences
Swapnil Hingmire, TCS Research, India
Tadashi Nomoto, National Institute of Japanese Literature
Tanik Saikh, IIT Patna, India
Tanmoy Chakraborty, IIIT Delhi India
Tapas Nayak, TCS Research and Innovation
Thiyam Susma Devi, IIT Guwahati, India
Tollef Jørgensen, NTNU (DART, IDI)
Vikas Verma, KCL IMT Jalandhar
Wes Robbins, Montana State University
Ye Liu, Mercedes-Benz AG, Sindelfingen, Germany
Zhongkai Sun, Amazon Alexa AI

# Organized by

National Institute of Technology Silchar

Natural Language Processing Association of India

# Sponsored by

SAMSUNG

Samsung

# Table of Contents

*Constrained Decoding for Technical Term Retention in English-Hindi MT*
Niyati Bafna, Martin Vastl and Ondřej Bojar . . . . . 1

*Named Entity-Factored Transformer for Proper Noun Translation*
Kohichi Takai, Gen Hattori, Akio Yoneyama, Keiji Yasuda, Katsuhito Sudoh and Satoshi Nakamura 7

*Multi-Task Learning for Improving Gender Accuracy in Neural Machine Translation*
Carlos Escolano, Graciela Ojeda, Christine Basta and Marta R. Costa-jussa . . . . . 12

*Small Batch Sizes Improve Training of Low-Resource Neural MT*
Àlex Atrio and Andrei Popescu-Belis . . . . . 18

*lakṣyārtha (Indicated Meaning) of Śabdavyāpāra (Function of a Word) framework from kāvyaśāstra (The Science of Literary Studies) in Samskṛtam : Its application to Literary Machine Translation and other NLP tasks*
Sripathi Sripada, Anupama Ryali and Raghuram Sheshadri . . . . . 25

*EduMT: Developing Machine Translation System for Educational Content in Indian Languages*
Ramakrishna Appicharla, Asif Ekbal and Pushpak Bhattacharyya . . . . . 35

*Assessing Post-editing Effort in the English-Hindi Direction*
Arafat Ahsan, Vandan Mujadia and Dipti Misra Sharma . . . . . 44

*An Experiment on Speech-to-Text Translation Systems for Manipuri to English on Low Resource Setting*
Loitongbam Sanayai Meetei, Laishram Rahul, Alok Singh, Salam Michael Singh, Thoudam Doren Singh and Sivaji Bandyopadhyay . . . . . 54

*On the Transferability of Massively Multilingual Pretrained Models in the Pretext of the Indo-Aryan and Tibeto-Burman Languages*
Salam Michael Singh, Loitongbam Sanayai Meetei, Alok Singh, Thoudam Doren Singh and Sivaji Bandyopadhyay . . . . . 64

*Generating Slogans with Linguistic Features using Sequence-to-Sequence Transformer*
Yeoun Yi and Hyopil Shin . . . . . 75

*Using Integrated Gradients and Constituency Parse Trees to explain Linguistic Acceptability learnt by BERT*
Anmol Nayak and Hari Prasad Timmapathini . . . . . 80

*The Importance of Context in Very Low Resource Language Modeling*
Lukas Edman, Antonio Toral and Gertjan van Noord . . . . . 86

*Stylistic MR-to-Text Generation Using Pre-trained Language Models*
Kunal Pagarey, Kanika Kalra, Abhay Garg, Saumajit Saha, Mayur Patidar and Shirish Karande . 93

*Deep Learning Based Approach For Detecting Suicidal Ideation in Hindi-English Code-Mixed Text: Baseline and Corpus*
Kaustubh Agarwal and Bhavya Dhingra . . . . . 100

*On the Universality of Deep Contextual Language Models*
Shaily Bhatt, Poonam Goyal, Sandipan Dandapat, Monojit Choudhury and Sunayana Sitaram . 106

*Towards Explainable Dialogue System: Explaining Intent Classification using Saliency Techniques*
Ratnesh Joshi, Arindam Chatterjee and Asif Ekbal . . . . . 120

*Comparing in context: Improving cosine similarity measures with a metric tensor*
Isa M. Apallius de Vos, Ghislaine L. van den Boogerd, Mara D. Fennema and Adriana Correia 128

*Context Matters in Semantically Controlled Language Generation for Task-oriented Dialogue Systems*
Ye Liu, Wolfgang Maier, Wolfgang Minker and Stefan Ultes . . . . . 139

*Data Augmentation for Mental Health Classification on Social Media*
Gunjan Ansari, Muskan Garg and Chandni Saxena . . . . . 152

*VAE based Text Style Transfer with Pivot Words Enhancement Learning*
Haoran Xu, Sixing Lu, Zhongkai Sun, Chengyuan Ma and Chenlei Guo . . . . . 162

*MRE : Multi Relationship Extractor for Persona based Empathetic Conversational Model*
Bharatram Natarajan and Abhijit Nargund . . . . . 173

*An End-to-End Speech Recognition for the Nepali Language*
Sunil Regmi and Bal Krishna Bal . . . . . 180

*Impact of Microphone position Measurement Error on Multi Channel Distant Speech Recognition & Intelligibility*
Karan Nathwani and Sunil Kumar Kopparapu . . . . . 186

*IE-CPS Lexicon: An Automatic Speech Recognition Oriented Indian-English Pronunciation Dictionary*
Shelly Jain, Aditya Yadavalli, Ganesh Mirishkar, Chiranjeevi Yarra and Anil Kumar Vuppala . 195

*An Investigation of Hybrid architectures for Low Resource Multilingual Speech Recognition system in Indian context*
Ganesh Mirishkar, Aditya Yadavalli and Anil Kumar Vuppala . . . . . 205

*Improve Sinhala Speech Recognition Through e2e LF-MMI Model*
Buddhi Gamage, Randil Pushpananda, Thilini Nadungodage and Ruwan Weerasinghe . . . . . 213

*Towards Multimodal Vision-Language Models Generating Non-Generic Text*
Wes Robbins, Zanyar Zohourianshahzadi and Jugal Kalita . . . . . 220

*Image Caption Generation Framework for Assamese News using Attention Mechanism*
Ringki Das and Thoudam Doren Singh . . . . . 231

*An Efficient Keyframes Selection Based Framework for Video Captioning*
Alok Singh, Loitongbam Sanayai Meetei, Salam Michael Singh, Thoudam Doren Singh and Sivaji Bandyopadhyay . . . . . 240

*A Scaled Encoder Decoder Network for Image Captioning in Hindi*
Santosh Kumar Mishra, Sriparna Saha and Pushpak Bhattacharyya . . . . . 251

*Co-attention based Multimodal Factorized Bilinear Pooling for Internet Memes Analysis*
Gitanjali Kumari, Amitava Das and Asif Ekbal . . . . . 261

*How effective is incongruity? Implications for code-mixed sarcasm detection*
Aditya Shah and Chandresh Maurya . . . . . 271

*Contrastive Learning of Sentence Representations*
Hefei Qiu, Wei Ding and Ping Chen . . . . . 277

*Classifying Verses of the Quran using Doc2vec*
Menwa Alshammeri, Eric Atwell and Mohammad Alsalka . . . . . 284

*ABB-BERT: A BERT model for disambiguating abbreviations and contractions*
Prateek Kacker, Andi Cupallari, Aswin Subramanian and Nimit Jain . . . . . 289

*Training data reduction for multilingual Spoken Language Understanding systems*
Anmol Bansal, Anjali Shenoy, Krishna Chaitanya Pappu, Kay Rottmann and Anurag Dwarakanath 298

*Leveraging Expectation Maximization for Identifying Claims in Low Resource Indian Languages*
Rudra Dhar and Dipankar Das . . . . . 307

*Performance of BERT on Persuasion for Good*
Saumajit Saha, Kanika Kalra, Manasi Patwardhan and Shirish Karande . . . . . 313

*Multi-Turn Target-Guided Topic Prediction with Monte Carlo Tree Search*
Jingxuan Yang, Si Li and Jun Guo . . . . . 324

*Resolving Prepositional Phrase Attachment Ambiguities with Contextualized Word Embeddings*
Adwait Ratnaparkhi and Atul Kumar . . . . . 335

*Multi-Source Cross-Lingual Constituency Parsing*
Hour Kaing, Chenchen Ding, Katsuhito Sudoh, Masao Utiyama, Eiichiro Sumita and Satoshi Nakamura . . . . . 341

*Kannada Sandhi Generator for Lopa and Adesha Sandhi*
Musica Supriya, Dinesh U. Acharya, Ashalatha Nayak and Arjuna S. R . . . . . 347

*Data Augmentation for Low-Resource Named Entity Recognition Using Backtranslation*
Usama Yaseen and Stefan Langer . . . . . 352

*Semantics of Spatio-Directional Geometric Terms of Indian Languages*
Sukhada Sukhada, Paul Soma, Rahul Kumar and Karthik Puranik . . . . . 359

*Morpheme boundary Detection & Grammatical feature Prediction for Gujarati : Dataset & Model*
Jatayu Baxi and Brijesh Bhatt . . . . . 369

*Auditing Keyword Queries Over Text Documents*
Bharath Kumar Reddy Apparreddy, Sailaja Rajanala and Manish Singh . . . . . 378

*A Method to Disambiguate a Word by Using Restricted Boltzmann Machine*
Nazreena Rahman and Bhogeswar Borah . . . . . 388

*Encoder Decoder Approach to Automated Essay Scoring For Deeper Semantic Analysis*
Priyatam Naravajhula, Sreedeep Rayavarapu and Srujana Inturi . . . . . 399

*Temporal Question Generation from History Text*
Harsimran Bedi, Sangameshwar Patil and Girish Palshikar . . . . . 408

*CAWESumm: A Contextual and Anonymous Walk Embedding Based Extractive Summarization of Legal Bills*
Deepali Jain, Malaya Dutta Borah and Anupam Biswas . . . . . 414

*Multi-document Text Summarization using Semantic Word and Sentence Similarity: A Combined Approach*
Rajendra Roul . . . . . 423

*#covid is war and #vaccine is weapon? COVID-19 metaphors in India*
Mohammed Khaliq, Rohan Joseph and Sunny Rai . . . . . 431

*Studies Towards Language Independent Fake News Detection*
Soumayan Majumder and Dipankar Das . . . . . 439

*Wikipedia Current Events Summarization using Particle Swarm Optimization*
Santosh Kumar Mishra, Darsh Kaushik, Sriparna Saha and Pushpak Bhattacharyya . . . . . 447

*Automated Evidence Collection for Fake News Detection*
Mrinal Rawat and Diptesh Kanojia . . . . . 456

*Prediction of Video Game Development Problems Based on Postmortems using Different Word Embedding Techniques*
Anirudh A, Aman RAJ Singh, Anjali Goyal, Lov Kumar and N L Bhanu Murthy . . . . . 465

*Multi-task pre-finetuning for zero-shot cross lingual transfer*
Moukthika Yerramilli, Pritam Varma and Anurag Dwarakanath . . . . . 474

*Sentiment Analysis For Bengali Using Transformer Based Models*
Anirban Bhowmick and Abhik Jana . . . . . 481

*IndicFed: A Federated Approach for Sentiment Analysis in Indic Languages*
Jash Mehta, Deep Gandhi, Naitik Rathod and Sudhir Bagul . . . . . 487

*An Efficient BERT Based Approach to Detect Aggression and Misogyny*
Sandip Dutta, Utso Majumder and Sudip Naskar . . . . . 493

*How vulnerable are you? A Novel Computational Psycholinguistic Analysis for Phishing Influence Detection*
Anik Chatterjee and Sagnik Basu . . . . . 499

*Aspect Based Sentiment Analysis Using Spectral Temporal Graph Neural Network*
Abir Chakraborty . . . . . 508

*Using Random Perturbations to Mitigate Adversarial Attacks on Sentiment Analysis Models*
Abigail Swenor and Jugal Kalita . . . . . 519

*Retrofitting of Pre-trained Emotion Words with VAD-dimensions and the Plutchik Emotions*
Manasi Kulkarni and Pushpak Bhattacharyya . . . . . 529

*Evaluating Pretrained Transformer Models for Entity Linking inTask-Oriented Dialog*
Sai Muralidhar Jayanthi, Varsha Embar and Karthik Raghunathan . . . . . 537

*Cascading Adaptors to Leverage English Data to Improve Performance of Question Answering for Low-Resource Languages*
Hariom Pandya, Bhavik Ardeshna and Brijesh Bhatt . . . . . 544

*eaVQA: An Experimental Analysis on Visual Question Answering Models*
Souvik Chowdhury and Badal Soni . . . . . 550

*Deep Embedding of Conversation Segments*
Abir Chakraborty and Anirban Majumder . . . . . 555

*DialogActs based Search and Retrieval for Response Generation in Conversation Systems*
Nidhi Arora, Rashmi Prasad and Srinivas Bangalore . . . . . 564

*An On-device Deep-Learning Approach for Attribute Extraction from Heterogeneous Unstructured Text*
Mahesh Gorijala, Aniruddha Bala, Pinaki Bhaskar, Krishnaditya . and Vikram Mupparthi . . . . . 573

*Weakly Supervised Extraction of Tasks from Text*
Sachin Pawar, Girish Palshikar and Anindita Sinha Banerjee . . . . . 583

*A German Corpus of Reflective Sentences*
Veronika Solopova, Oana-Iuliana Popescu, Margarita Chikobava, Ralf Romeike, Tim Landgraf and Christoph Benzmüller . . . . . 593

*Analysis of Manipuri Tones in ManiTo: A Tonal Contrast Database*
Thiyam Susma Devi and Pradip K. Das . . . . . 601

*Building a Linguistic Resource : A Word Frequency List for Sinhala*
Aloka Fernando and Gihan Dias . . . . . 606

*Part of Speech Tagging for a Resource Poor Language : Sindhi in Devanagari Script using HMM and CRF*
Bharti Nathani and Nisheeth Joshi . . . . . 611

*Stress Rules from Surface Forms: Experiments with Program Synthesis*
Saujas Vaduguru, Partho Sarthi, Monojit Choudhury and Dipti Sharma . . . . . 619

*Cross-lingual Alignment of Knowledge Graph Triples with Sentences*
Swayatta Daw, Shivprasad Sagare, Tushar Abhishek, Vikram Pudi and Vasudeva Varma . . . . . 629

*Introduction to ProverbNet: An Online Multilingual Database of Proverbs and Comprehensive Metadata*
Shreyas Pimpalgaonkar, Dhanashree Lele, Malhar Kulkarni and Pushpak Bhattacharyya . . . . . 638

*Bypassing Optimization Complexity through Transfer Learning & Deep Neural Nets for Speech Intelligibility Improvement*
Ritujoy Biswas . . . . . 651

*Design and Development of Spoken Dialogue System in Indic Languages*
Shrikant Malviya . . . . . 654

*FinRead: A Transfer Learning Based Tool to Assess Readability of Definitions of Financial Terms*
Sohom Ghosh, Shovon Sengupta, Sudip Naskar and Sunny Kumar Singh . . . . . 658

*Demo of the Linguistic Field Data Management and Analysis System - LiFE*
Siddharth Singh, Ritesh Kumar, Shyam Ratan and Sonal Sinha . . . . . 660

*Text Based Smart Answering System in Agriculture using RNN*
Raji Sukumar, Hemalatha N, Sarin S and Rose Mary C A . . . . . 663

*Image2tweet: Datasets in Hindi and English for Generating Tweets from Images*
Rishabh Jha, Varshith Kaki, Varuna Kolla, Shubham Bhagat, Parth Patwa, Amitava Das and Santanu Pal . . . . . 670

# Conference Program

**Day 1: Friday, December 17, 2021**

10:00 AM–11:00 AM **Inaugural Ceremony**

11:30 AM–12:30 PM **Keynote Lecture 1:** Phrases and Self-Supervision in Neural Machine Translation
Prof. Josef van Genabith
**Session Chair:** Sivaji Bandyopadhyay

12:30 PM–02:30 PM **BREAK**

02:30PM–06:00PM **Technical Session I:** Machine Translation
**Session Chair:** Partha Pakray **Co-Chair:** Badal Soni

*Constrained Decoding for Technical Term Retention in English-Hindi MT*
Niyati Bafna, Martin Vastl and Ondřej Bojar

*Named Entity-Factored Transformer for Proper Noun Translation*
Kohichi Takai, Gen Hattori, Akio Yoneyama, Keiji Yasuda, Katsuhito Sudoh and Satoshi Nakamura

*Multi-Task Learning for Improving Gender Accuracy in Neural Machine Translation*
Carlos Escolano, Graciela Ojeda, Christine Basta and Marta R. Costa-jussa

*Small Batch Sizes Improve Training of Low-Resource Neural MT*
Àlex Atrio and Andrei Popescu-Belis

*lakṣyārtha (Indicated Meaning) of Śabdavyāpāra (Function of a Word) framework from kāvyaśāstra (The Science of Literary Studies) in Samskṛtam : Its application to Literary Machine Translation and other NLP tasks*
Sripathi Sripada, Anupama Ryali and Raghuram Sheshadri

*EduMT: Developing Machine Translation System for Educational Content in Indian Languages*
Ramakrishna Appicharla, Asif Ekbal and Pushpak Bhattacharyya

*Assessing Post-editing Effort in the English-Hindi Direction*
Arafat Ahsan, Vandan Mujadia and Dipti Misra Sharma

*An Experiment on Speech-to-Text Translation Systems for Manipuri to English on Low Resource Setting*
Loitongbam Sanayai Meetei, Laishram Rahul, Alok Singh, Salam Michael Singh, Thoudam Doren Singh and Sivaji Bandyopadhyay

**Day 1: Friday, December 17, 2021 (continued)**

*On the Transferability of Massively Multilingual Pretrained Models in the Pretext of the Indo-Aryan and Tibeto-Burman Languages*
Salam Michael Singh, Loitongbam Sanayai Meetei, Alok Singh, Thoudam Doren Singh and Sivaji Bandyopadhyay

02:30 PM–06:00 PM **Technical Session II:** Natural Language Text Generation
**Session Chair:** Vasudeva Varma **Co-Chair:** Pinki Roy

*Generating Slogans with Linguistic Features using Sequence-to-Sequence Transformer*
Yeoun Yi and Hyopil Shin

*Using Integrated Gradients and Constituency Parse Trees to explain Linguistic Acceptability learnt by BERT*
Anmol Nayak and Hari Prasad Timmapathini

*The Importance of Context in Very Low Resource Language Modeling*
Lukas Edman, Antonio Toral and Gertjan van Noord

*Stylistic MR-to-Text Generation Using Pre-trained Language Models*
Kunal Pagarey, Kanika Kalra, Abhay Garg, Saumajit Saha, Mayur Patidar and Shirish Karande

*Deep Learning Based Approach For Detecting Suicidal Ideation in Hindi-English Code-Mixed Text: Baseline and Corpus*
Kaustubh Agarwal and Bhavya Dhingra

*On the Universality of Deep Contextual Language Models*
Shaily Bhatt, Poonam Goyal, Sandipan Dandapat, Monojit Choudhury and Sunayana Sitaram

*Towards Explainable Dialogue System: Explaining Intent Classification using Saliency Techniques*
Ratnesh Joshi, Arindam Chatterjee and Asif Ekbal

*Comparing in context: Improving cosine similarity measures with a metric tensor*
Isa M. Apallius de Vos, Ghislaine L. van den Boogerd, Mara D. Fennema and Adriana Correia

*Context Matters in Semantically Controlled Language Generation for Task-oriented Dialogue Systems*
Ye Liu, Wolfgang Maier, Wolfgang Minker and Stefan Ultes

*Data Augmentation for Mental Health Classification on Social Media*
Gunjan Ansari, Muskan Garg and Chandni Saxena

**Day 1: Friday, December 17, 2021 (continued)**

*VAE based Text Style Transfer with Pivot Words Enhancement Learning*
Haoran Xu, Sixing Lu, Zhongkai Sun, Chengyuan Ma and Chenlei Guo

*MRE : Multi Relationship Extractor for Persona based Empathetic Conversational Model*
Bharatram Natarajan and Abhijit Nargund

02:30 PM–06:00 PM **Technical Session III:** Speech / Multimodality
**Session Chair:** Kalika Bali **Co-Chair:** Rabul Hussain Laskar

*An End-to-End Speech Recognition for the Nepali Language*
Sunil Regmi and Bal Krishna Bal

*Impact of Microphone position Measurement Error on Multi Channel Distant Speech Recognition & Intelligibility*
Karan Nathwani and Sunil Kumar Kopparapu

*IE-CPS Lexicon: An Automatic Speech Recognition Oriented Indian-English Pronunciation Dictionary*
Shelly Jain, Aditya Yadavalli, Ganesh Mirishkar, Chiranjeevi Yarra and Anil Kumar Vuppala

*An Investigation of Hybrid architectures for Low Resource Multilingual Speech Recognition system in Indian context*
Ganesh Mirishkar, Aditya Yadavalli and Anil Kumar Vuppala

*Improve Sinhala Speech Recognition Through e2e LF-MMI Model*
Buddhi Gamage, Randil Pushpananda, Thilini Nadungodage and Ruwan Weerasinghe

*Towards Multimodal Vision-Language Models Generating Non-Generic Text*
Wes Robbins, Zanyar Zohourianshahzadi and Jugal Kalita

*Image Caption Generation Framework for Assamese News using Attention Mechanism*
Ringki Das and Thoudam Doren Singh

*An Efficient Keyframes Selection Based Framework for Video Captioning*
Alok Singh, Loitongbam Sanayai Meetei, Salam Michael Singh, Thoudam Doren Singh and Sivaji Bandyopadhyay

*A Scaled Encoder Decoder Network for Image Captioning in Hindi*
Santosh Kumar Mishra, Sriparna Saha and Pushpak Bhattacharyya

**Day 1: Friday, December 17, 2021 (continued)**

*Co-attention based Multimodal Factorized Bilinear Pooling for Internet Memes Analysis*
Gitanjali Kumari, Amitava Das and Asif Ekbal

06:00 PM–06:30 PM **TEA BREAK**

06:30 PM–07:30 PM **Keynote Lecture 2:** Mental Health as an Application Area for Natural Language Processing: Prospects and Challenges
Prof. Philip Resnik
**Session Chair:** Dipti Misra Sharma

07:30 PM–07:45 PM **BUFFER**

07:45 PM–08:45 PM **Keynote Lecture 3:** Multimodal AI: Understanding Human Behaviors
Dr. Louis-Philippe Morency
**Session Chair:** Sivaji Bandyopadhyay

**Day 2: Saturday, December 18, 2021**

9:00 AM–12:00 PM **Technical Session IV:** Machine Learning in NLP
**Session Chair:** Sudeshna Sarkar **Co-Chair:** Anupam Biswas

*How effective is incongruity? Implications for code-mixed sarcasm detection*
Aditya Shah and Chandresh Maurya

*Contrastive Learning of Sentence Representations*
Hefei Qiu, Wei Ding and Ping Chen

*Classifying Verses of the Quran using Doc2vec*
Menwa Alshammeri, Eric Atwell and Mohammad Alsalka

*ABB-BERT: A BERT model for disambiguating abbreviations and contractions*
Prateek Kacker, Andi Cupallari, Aswin Subramanian and Nimit Jain

*Training data reduction for multilingual Spoken Language Understanding systems*
Anmol Bansal, Anjali Shenoy, Krishna Chaitanya Pappu, Kay Rottmann and Anurag Dwarakanath

*Leveraging Expectation Maximization for Identifying Claims in Low Resource Indian Languages*
Rudra Dhar and Dipankar Das

**Day 2: Saturday, December 18, 2021 (continued)**

*Performance of BERT on Persuasion for Good*
Saumajit Saha, Kanika Kalra, Manasi Patwardhan and Shirish Karande

*Multi-Turn Target-Guided Topic Prediction with Monte Carlo Tree Search*
Jingxuan Yang, Si Li and Jun Guo

9:00 AM–12:00 PM **Technical Session V:** Syntax and Semantics
**Session Chair:** Dipti Misra Sharma **Co-Chair:** Samir Kumar Borgohain

*Resolving Prepositional Phrase Attachment Ambiguities with Contextualized Word Embeddings*
Adwait Ratnaparkhi and Atul Kumar

*Multi-Source Cross-Lingual Constituency Parsing*
Hour Kaing, Chenchen Ding, Katsuhito Sudoh, Masao Utiyama, Eiichiro Sumita and Satoshi Nakamura

*Kannada Sandhi Generator for Lopa and Adesha Sandhi*
Musica Supriya, Dinesh U. Acharya, Ashalatha Nayak and Arjuna S. R

*Data Augmentation for Low-Resource Named Entity Recognition Using Backtranslation*
Usama Yaseen and Stefan Langer

*Semantics of Spatio-Directional Geometric Terms of Indian Languages*
Sukhada Sukhada, Paul Soma, Rahul Kumar and Karthik Puranik

*Morpheme boundary Detection & Grammatical feature Prediction for Gujarati : Dataset & Model*
Jatayu Baxi and Brijesh Bhatt

*Auditing Keyword Queries Over Text Documents*
Bharath Kumar Reddy Apparreddy, Sailaja Rajanala and Manish Singh

*A Method to Disambiguate a Word by Using Restricted Boltzmann Machine*
Nazreena Rahman and Bhogeswar Borah

*Encoder Decoder Approach to Automated Essay Scoring For Deeper Semantic Analysis*
Priyatam Naravajhula, Sreedeep Rayavarapu and Srujana Inturi

**Day 2: Saturday, December 18, 2021 (continued)**

9:00 AM–12:00 PM **Technical Session VI:** NLP Applications
**Session Chair:** Amitava Das **Co-Chair:** Sajor Kumar Biswas

*Temporal Question Generation from History Text*
Harsimran Bedi, Sangameshwar Patil and Girish Palshikar

*CAWESumm: A Contextual and Anonymous Walk Embedding Based Extractive Summarization of Legal Bills*
Deepali Jain, Malaya Dutta Borah and Anupam Biswas

*Multi-document Text Summarization using Semantic Word and Sentence Similarity: A Combined Approach*
Rajendra Roul

*#covid is war and #vaccine is weapon? COVID-19 metaphors in India*
Mohammed Khaliq, Rohan Joseph and Sunny Rai

*Studies Towards Language Independent Fake News Detection*
Soumayan Majumder and Dipankar Das

*Wikipedia Current Events Summarization using Particle Swarm Optimization*
Santosh Kumar Mishra, Darsh Kaushik, Sriparna Saha and Pushpak Bhattacharyya

*Automated Evidence Collection for Fake News Detection*
Mrinal Rawat and Diptesh Kanojia

*Prediction of Video Game Development Problems Based on Postmortems using Different Word Embedding Techniques*
Anirudh A, Aman RAJ Singh, Anjali Goyal, Lov Kumar and N L Bhanu Murthy

*Multi-task pre-finetuning for zero-shot cross lingual transfer*
Moukthika Yerramilli, Pritam Varma and Anurag Dwarakanath

01:30 PM–02:30 PM **BREAK**

02:30 PM–03:30 PM **Keynote Lecture 4:** Human-centered Natural Language Processing
Prof. Rada Mihalcea
**Session Chair:** Rajeev Sangal

**Day 2: Saturday, December 18, 2021 (continued)**

03:30 PM–03:45 PM **BREAK**

03:45 PM–06:00 PM **Technical Session VII:** Computational Psycholinguistics/Sentiment Analysis and Emotion Recognition
**Session Chair:** Dipankar Das **Co-Chair:** Ujwala Baruah

*Sentiment Analysis For Bengali Using Transformer Based Models*
Anirban Bhowmick and Abhik Jana

*IndicFed: A Federated Approach for Sentiment Analysis in Indic Languages*
Jash Mehta, Deep Gandhi, Naitik Rathod and Sudhir Bagul

*An Efficient BERT Based Approach to Detect Aggression and Misogyny*
Sandip Dutta, Utso Majumder and Sudip Naskar

*How vulnerable are you? A Novel Computational Psycholinguistic Analysis for Phishing Influence Detection*
Anik Chatterjee and Sagnik Basu

*Aspect Based Sentiment Analysis Using Spectral Temporal Graph Neural Network*
Abir Chakraborty

*Using Random Perturbations to Mitigate Adversarial Attacks on Sentiment Analysis Models*
Abigail Swenor and Jugal Kalita

*Retrofitting of Pre-trained Emotion Words with VAD-dimensions and the Plutchik Emotions*
Manasi Kulkarni and Pushpak Bhattacharyya

**Day 2: Saturday, December 18, 2021 (continued)**

03:45 PM–06:00 PM **Technical Session VIII:** QA/Information Extraction/Information Retrieval/Text Mining
**Session Chair:** Ranjani Parthasarathi **Co-Chair:** Partha Pakray

*Evaluating Pretrained Transformer Models for Entity Linking inTask-Oriented Dialog*
Sai Muralidhar Jayanthi, Varsha Embar and Karthik Raghunathan

*Cascading Adaptors to Leverage English Data to Improve Performance of Question Answering for Low-Resource Languages*
Hariom Pandya, Bhavik Ardeshna and Brijesh Bhatt

*eaVQA: An Experimental Analysis on Visual Question Answering Models*
Souvik Chowdhury and Badal Soni

*Deep Embedding of Conversation Segments*
Abir Chakraborty and Anirban Majumder

*DialogActs based Search and Retrieval for Response Generation in Conversation Systems*
Nidhi Arora, Rashmi Prasad and Srinivas Bangalore

*An On-device Deep-Learning Approach for Attribute Extraction from Heterogeneous Unstructured Text*
Mahesh Gorijala, Aniruddha Bala, Pinaki Bhaskar, Krishnaditya . and Vikram Muppparthi

*Weakly Supervised Extraction of Tasks from Text*
Sachin Pawar, Girish Palshikar and Anindita Sinha Banerjee

**Day 2: Saturday, December 18, 2021 (continued)**

03:45 PM–06:00 PM **Technical Session IX:** Language Resources and Evaluation
**Session Chair:** Girish Nath Jha **Co-Chair:** Shyamapada Mukherjee

*A German Corpus of Reflective Sentences*
Veronika Solopova, Oana-Iuliana Popescu, Margarita Chikobava, Ralf Romeike, Tim Landgraf and Christoph Benzmüller

*Analysis of Manipuri Tones in ManiTo: A Tonal Contrast Database*
Thiyam Susma Devi and Pradip K. Das

*Building a Linguistic Resource : A Word Frequency List for Sinhala*
Aloka Fernando and Gihan Dias

*Part of Speech Tagging for a Resource Poor Language : Sindhi in Devanagari Script using HMM and CRF*
Bharti Nathani and Nisheeth Joshi

*Stress Rules from Surface Forms: Experiments with Program Synthesis*
Saujas Vaduguru, Partho Sarthi, Monojit Choudhury and Dipti Sharma

*Cross-lingual Alignment of Knowledge Graph Triples with Sentences*
Swayatta Daw, Shivprasad Sagare, Tushar Abhishek, Vikram Pudi and Vasudeva Varma

*Introduction to ProverbNet: An Online Multilingual Database of Proverbs and Comprehensive Metadata*
Shreyas Pimpalgaonkar, Dhanashree Lele, Malhar Kulkarni and Pushpak Bhattacharyya

06:00 PM–07:00 PM **Valedictory Session**

07:00 PM–08:00 PM **NLPAI Meeting**

**Pre-conference: Thursday, December 16, 2021**

10:00 AM – 12:30 PM **Doctoral Consortium Session**

**Session Chair:** Radhika Mamidi **Co-Chair:** Samar Hussain

*Bypassing Optimization Complexity through Transfer Learning & Deep Neural Nets for Speech Intelligibility Improvement*
Ritujoy Biswas

*Design and Development of Spoken Dialogue System in Indic Languages*
Shrikant Malviya

**Post-Conference: Sunday, December 19, 2021**

10:00AM – 05:00 PM **Tools, Demo and Shared Tasks**
**Chair:** Partha Pakray **Co-Chair:** Vishal Goyal

*FinRead: A Transfer Learning Based Tool to Assess Readability of Definitions of Financial Terms*
Sohom Ghosh, Shovon Sengupta, Sudip Naskar and Sunny Kumar Singh

*Demo of the Linguistic Field Data Management and Analysis System - LiFE*
Siddharth Singh, Ritesh Kumar, Shyam Ratan and Sonal Sinha

*Text Based Smart Answering System in Agriculture using RNN*
Raji Sukumar, Hemalatha N, Sarin S and Rose Mary C A

*Image2tweet: Datasets in Hindi and English for Generating Tweets from Images*
Rishabh Jha, Varshith Kaki, Varuna Krishna Kolla, Shubham Bhagat, Parth Patwa, Amitava Das and Santanu Pal