

ICON 2019

**16th International
Conference on Natural
Language Processing**

Proceedings of the Conference

18-21 December 2019
International Institute of Information Technology, Hyderabad,
India

©2019 NLP Association of India (NLPAI)

Preface

Research in Natural Language Processing (NLP) has taken a noticeable leap in the recent years. 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, industry is keenly interested in obtaining NLP technology for mass use. The 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 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 which 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. Larger set of corpora are expected to be prepared in the near future.

This conference proceedings contains papers selected for presentation in technical sessions of ICON-2019 and short communications selected for poster presentation. 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 88 submissions, 28 papers were selected, 19 for full presentation, 9 for poster presentation, representing a variety of new and interesting developments, covering a wide spectrum of NLP areas and core linguistics. Besides presentations, the conference also hosted 5 tutorials and 1 workshop.

We are deeply grateful to Jan Hajič from Charles University (Czech Republic), C. V. Jawahar from IIIT Hyderabad (India) and Amba Kulkarni from University of Hyderabad (India) for giving the keynote lectures at ICON. We would also like to thank the members of the Advisory Committee and Programme Committee for their support and cooperation in making ICON 2019 a success.

We thank Anil Kumar Singh, Samar Hussain, Co-Chairs, Workshop/ Tutorial for taking the responsibilities of the events. We are thankful to Radhika Mamidi, Anil Kumar Vuppala and Manish Shrivastava for making the organization of the event at the International Institute of Information Technology (IIIT) a success.

We convey our thanks to Krishna Sireesha S., P V S Ram Babu, G Srinivas Rao, A Lakshmi Narayana, Praveen, Hostel and Administrative staff, International Institute of Information Technology (IIIT), Hyderabad for their dedicated efforts in successfully handling the ICON Secretariat. We also thank IIIT Hyderabad team of Vineet Chaitanya, Vasudeva Varma, Soma Paul, Radhika Mamidi, Manish Shrivastava, Suryakanth V Gangashetty and Anil Kumar Vuppala. We heartily express our gratitude to Saumitra Yadav, Pruthwik Mishra, Vandan Mujadia, Nirmal Surange, Prashant Kodali and other team members at IIIT Hyderabad for their timely help with sincere dedication to make this conference a success. We also thank all those who came forward to help us in this task.

Finally, we thank all the researchers who responded to our call for papers and all the participants of ICON-2019, without whose overwhelming response the conference would not have been a success.

December 2019
Hyderabad

Pushpak Bhattacharyya
Dipti Misra Sharma
Rajeev Sangal

Conference General Chair :

Rajeev Sangal, IIIT Hyderabad, India

Program Committee:

Dipti Misra Sharma, IIIT Hyderabad, India (Co-Chair)
Pushpak Bhattacharyya, IIT Bombay and Patna, India (Co-Chair)
Balamurali A R , LIF AIX CNRS
Bharat Ram Ambati , Apple Inc.
Rakesh Balabantaray , IIIT Bhubaneswar
Monojit Choudhury , Microsoft Research
Dipankar Das , Jadavpur University
Amitava Das , Wipro AI Lab
Asif Ekbal , IIT Patna
Sanjukta Ghosh , IIT (BHU)
Vishal Goyal , Punjabi University Patiala
Pawan Goyal , IIT Kharagpur
Harald Hammarström , Uppsala University
Samar Husain , IIT Delhi
C V Jawahar , IIIT Hyderabad
Aditya Joshi , CSIRO
Mitesh M. Khapra , IIT Madras
Parameswari Krishnamurthy , University of Hyderabad
Amba Kulkarni , University of Hyderabad
Malhar Kulkarni , IIT Bombay
Ritesh Kumar , Dr. Bhimrao Ambedkar University, Agra
Niraj Kumar , Conduent Labs India
Anil Kumar Singh , IIT (BHU), Varanasi
Anoop Kunchukuttan , Microsoft AI and Research
Bornini Lahiri , IIT Kharagpur
Sobha Lalitha Devi , Au-KBC Research Centre, Anna University
Gurpreet Lehal , Punjabi University
Radhika Mamidi , IIIT Hyderabad
Yuji Matsumoto , Nara Institute of Science and Technology
Pruthwik Mishra , IIIT, Hyderabad
Vinay Kumar Mittal , IIIT Chittoor, Sri City
Ashutosh Modi , IIT Kanpur
Aditya Mogadala , Saarland University
Vandan Mujadia , IIIT, Hyderabad
Animesh Mukherjee , IIT Kharagpur
Jai Nanavati , Navana Tech
Sudip Kumar Naskar , Jadavpur University
Vasudevan Nedumpozhi , TU Dublin
Deepak P , Queen's University Belfast

Girish Palshikar , Tata Consultancy Services Limited
Tanvina Patel , Cogknit Semantics
Dripta Piplai , IIT Kharagpur
Bapi Raju , IIIT, Hyderabad
Sai Krishna Rallabandi , Carnegie Mellon University
Paolo Rosso , Universitat Politècnica de València
Atanu Saha , Jadavpur University
Peter Scharf , IIIT, Hyderabad
Raksha Sharma , IIT Roorkee
Ravi Shekhar , Queen Mary University of London
Elizabeth Sherly , IIITM-Kerala
Manish Shrivastava , IIIT Hyderabad
Smriti Singh , Samsung Research UK
Manjira Sinha , IIT Kharagpur
Sunayana Sitaram , Microsoft Research India
Keh-Yih Su , Institute of Information Science, Academia Sinica
Partha Talukdar , Indian Institute of Science
Anil Thakur , IIT (BHU) Varanasi
Ashwini Vaidya , IIT Delhi
Sriram Venkatapathy , Amazon
Samudra Vijaya , IIT Guwahati
Anil Kumar Vuppala , IIIT Hyderabad
Saumitra Yadav , IIIT, Hyderabad

Organizing Committee:

Radhika Mamidi, IIIT Hyderabad
Manish Shrivastava, IIIT Hyderabad
Anil Kumar Vuppala, IIIT Hyderabad

Organized by:



Sponsored by:



Table of Contents

<i>Robust Text Classification using Sub-Word Information in Input Word Representations.</i> Bhanu Prakash Mahanti, Priyank Chhipa, Vivek Sridhar and Vinuthkumar Prasan	1
<i>A Deep Ensemble Framework for Fake News Detection and Multi-Class Classification of Short Political Statements</i> Arjun Roy, Kingshuk Basak, Asif Ekbal and Pushpak Bhattacharyya	9
<i>Building Discourse Parser for Thirukkural</i> Anita R and Subalalitha C N	18
<i>Introducing Aspects of Creativity in Automatic Poetry Generation</i> Brendan Bena and Jugal Kalita	26
<i>Incorporating Sub-Word Level Information in Language Invariant Neural Event Detection</i> Suhan Prabhu, Pranav Goel, Alok Debnath and Manish Shrivastava	36
<i>Event Centric Entity Linking for Hindi News Articles: A Knowledge Graph Based Approach</i> Pranav Goel, Suhan Prabhu, Alok Debnath and Manish Shrivastava	45
<i>Language Modelling with NMT Query Translation for Amharic-Arabic Cross-Language Information Retrieval</i> Ibrahim Gashaw and H.L Shashirekha	56
<i>Non-native Accent Partitioning for Speakers of Indian Regional Languages</i> RADHA KRISHNA GUNTUR, Krishnan Ramakrishnan and Vinay Kumar Mittal	65
<i>A little perturbation makes a difference: Treebank augmentation by perturbation improves transfer parsing</i> Ayan Das and Sudeshna Sarkar	75
<i>Autism Speech Analysis using Acoustic Features</i> Abhijit Mohanta and Vinay Kumar Mittal	85
<i>A Survey on Ontology Enrichment from Text</i> Vivek Iyer, Lalit Mohan, Mehar Bhatia and Y. Raghu Reddy	95
<i>Sanskrit Segmentation revisited</i> Sriram Krishnan and Amba Kulkarni	105
<i>Integrating Lexical Knowledge in Word Embeddings using Sprinkling and Retrofitting</i> Aakash Srinivasan, Harshavardhan Kamarthi, Devi Ganesan and Sutanu Chakraborti	115
<i>Robust Deep Learning Based Sentiment Classification of Code-Mixed Text</i> Siddhartha Mukherjee, Vinuthkumar Prasan, Anish Nediyanath, Manan Shah and Nikhil Kumar 124	
<i>Dataset for Aspect Detection on Mobile reviews in Hindi</i> Pruthwik Mishra, Ayush Joshi and Dipti Sharma	130
<i>Multi-linguality helps: Event-Argument Extraction for Disaster Domain in Cross-lingual and Multi-lingual setting</i> Zishan Ahmad, Deeksha Varshney, Asif Ekbal and Pushpak Bhattacharyya	135

<i>Development of POS tagger for English-Bengali Code-Mixed data</i> Tathagata Raha, Sainik Mahata, Dipankar Das and Sivaji Bandyopadhyay	143
<i>Towards Handling Verb Phrase Ellipsis in English-Hindi Machine Translation</i> Niyati Bafna and Dipti Sharma	150
<i>A Multi-task Model for Multilingual Trigger Detection and Classification</i> Sovan Kumar Sahoo, Saumajit Saha, Asif Ekbal and Pushpak Bhattacharyya	160
<i>Converting Sentiment Annotated Data to Emotion Annotated Data</i> Manasi Kulkarni and Pushpak Bhattacharyya	170
<i>Towards measuring lexical complexity in Malayalam</i> Richard Shallam and Ashwini Vaidya	178
<i>Kunji : A Resource Management System for Higher Productivity in Computer Aided Translation Tools</i> Priyank Gupta, Manish Shrivastava, Dipti Misra Sharma and Rashid Ahmad	184
<i>Identification of Synthetic Sentence in Bengali News using Hybrid Approach</i> Soma Das and Sanjay Chatterji	193
<i>An LSTM-Based Deep Learning Approach for Detecting Self-Deprecating Sarcasm in Textual Data</i> Ashraf Kamal and Muhammad Abulaish	201
<i>Unsung Challenges of Building and Deploying Language Technologies for Low Resource Language Communities</i> Pratik Joshi, Christain Barnes, Sebastin Santy, Simran Khanuja, Sanket Shah, Anirudh Srinivasan, Satwik Bhattamishra, Sunayana Sitaram, Monojit Choudhury and Kalika Bali	211
<i>DRCoVe: An Augmented Word Representation Approach using Distributional and Relational Context</i> Md. Aslam Parwez, Muhammad Abulaish and Mohd Fazil	220
<i>A Deep Learning Approach for Automatic Detection of Fake News</i> Tanik Saikh, Arkadipta De, Asif Ekbal and Pushpak Bhattacharyya	230
<i>Samajh-Boojh: A Reading Comprehension system in Hindi</i> Shalaka Vaidya, Hiranmai Sri Adibhatla and Radhika Mamidi	239

Conference Program

Day 1: Thursday, December 19, 2019

+ 9:00 – 9:30 Inaugural Ceremony

+ 9:30 – 10:30 Keynote I by Prof. Jan Hajič.

Title: *“European Language Technology and Resources Infrastructures”*

Session Chair: Prof. Rajeev Sangal

+ 10:30 – 11:00 Tea break

+ 11:00 – 12:30 **Technical Session I:** Information Retrieval and Extraction

Session Chair: Sobha L Devi

A Deep Ensemble Framework for Fake News Detection and Multi-Class Classification of Short Political Statements

Arjun Roy, Kingshuk Basak, Asif Ekbal and Pushpak Bhattacharyya

Language Modelling with NMT Query Translation for Amharic-Arabic Cross-Language Information Retrieval

Ibrahim Gashaw and H.L Shashirekha

Robust Deep Learning Based Sentiment Classification of Code-Mixed Text

Siddhartha Mukherjee, Vinuthkumar Prasan, Anish Nediyanath, Manan Shah, Nikhil Kumar and Prajakta Kulkarni

+ 12:30 – 13:30 Lunch

+ 13:30 – 14:30 Keynote II by Prof C. V. Jawahar

Title: *“Data Driven Methods for Multilingual and Multimodal Problems”*

Session Chair: Prof. Pushpak Bhattacharyya

+ 14:30 – 16:00

Technical Session II: Machine Learning in NLP

Session Chair: Asif Ekbal

Robust Text Classification using Sub-Word Information in Input Word Representations.

Bhanu Prakash, Priyank Chhipa, Vivek Sridhar and Vinuthkumar Prasan

Introducing Aspects of Creativity in Automatic Poetry Generation

Brendan Bena and Jugal Kalita

Identification of Synthetic Sentence in Bengali News using Hybrid Approach

Soma Das and Sanjay Chatterji

Technical Session III: Speech Processing
Session Chair: S. R. M. Prasanna

Non-native Accent Partitioning for Speakers of Indian Regional Languages
Guntur Radhakrishna, Krishnan Ramakrishnan and Vinay Kumar Mittal

Autism Speech Analysis using Acoustic Features
Abhijit Mohanta and Vinay Kumar Mittal

+ 16:00 – 17:00 Tea break

+ 16:00 – 17:00 Poster Session
Session Chair: Manish Shrivastava

+ 17:00 – 18:00 NLP AI Meeting

+ 18:30 – 20:00 Cultural Programme

+ 20:30 Onwards Dinner

Day 2: Friday, December 20, 2019

+ 9:00 – 10.30 Keynote-III by Prof. Amba Kulkarni

Title: “*Information Coding in Language: Some insights from Indian Grammatical Tradition*”

Session Chair: Sudeshna Sarkar

+ 10:30 – 11:00 Tea break

+ 11:00 – 13:00

Technical Session IV: Information Retrieval and Extraction

Session Chair: Ashutosh Modi

Multi-linguality helps: Event-Argument Extraction for Disaster Domain in Cross-lingual and Multilingual setting

Zishan Ahmad, Deeksha Varshney, Asif Ekbal and Pushpak Bhattacharyya

A Multi-task Model for Multilingual Trigger Detection and Classification

Sovan Kumar Sahoo, Saumajit Saha, Asif Ekbal and Pushpak Bhattacharyya

Unsung Challenges of Building and Deploying Language Technologies for Low Resource Language Communities

Pratik Joshi, Christain Barnes, Sebastin Santy, Simran Khanuja, Sanket Shah, Anirudh Srinivasan, Satwik Bhattamishra, Sunayana Sitaram, Monojit Choudhury and Kalika Bali

Samajh-Boojh: A Reading Comprehension system in Hindi

Shalaka Vaidya and Hiranmai Sri Adibhatla

Technical Session V: NLP for Low Resource Languages

Session Chair: Bharat Ambati

Incorporating Sub-Word Level Information in Language Invariant Neural Event Detection

Suhan Prabhu, Pranav Goel, Alok Debnath and Manish Shrivastava

A little perturbation makes a difference: Treebank augmentation by perturbation improves transfer parsing

Ayan Das and Sudeshna Sarkar

Sanskrit Segmentation revisited

Sriram Krishnan and Amba Kulkarni

Towards measuring lexical complexity in Malayalam

Richard Shallam and Ashwini Vaidya

+ 13:00 – 14:00 Lunch

+ 14:00 – 15:00 Women in NLP workshop

Session Chair: Manjira Sinha

+ 15:00 –15:30 Industry Interaction Session

+ 15:45 –17:00 **Technical Session VI:** Syntax, Semantics and Discourse

Session Chair: Radhika Mamidi

Building Discourse Parser for Thirukkural

Anita R and Subalalitha C N

DRCoVe: An Augmented Word Representation Approach using Distributional and Relational Context

Md. Aslam Parwez, Muhammad Abulaish and Mohd Fazil

Event Centric Entity Linking for Hindi News Articles: A Knowledge Graph Based Approach

Pranav Goel, Suhan Prabhu, Alok Debnath and Manish Shrivastava

+ 17:00 – 17:15 Tea break

+ 17:15 – 17:45 Valedictory Function