ICON-2016

13th International Conference on Natural Language Processing

Proceedings of the Conference

17-20 December 2016
IIT (BHU), Varanasi, India

© 2016 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 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 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 the 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 development of basic tools for multiple languages which are inter-operable for the tasks of machine translation, cross lingual search, hand writing 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.

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 near future.

This volume contains papers selected for presentation in technical sessions of ICON-2016 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 150 submissions, 38 papers were selected, 15 for full presentation, 18 for poster presentation and 5 for poster-cum-demonstration, representing a variety of new and interesting developments, covering a wide spectrum of NLP areas and core linguistics.

We are deeply grateful to Richard Sproat, Google Inc., USA and Bruno Pouliquen, World Intellectual Property Organization, Switzerland for giving the keynote lectures at ICON-2016. We would also like to thank the members of the Advisory Committee and Programme Committee for their support and co-operation in making ICON 2016 a success.

We thank Asif Ekbal, Chair, Student Paper Competition and Amitav Das, Chair, NLP Tools Contest for taking the responsibilities of the events.

We convey our thanks to P V S Ram Babu, G Srinivas Rao, B Mahender Kumar and A Lakshmi Narayana, 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, Peri Bhaskararao, Vasudeva Varma, Soma Paul, Radhika Mamidi, Manish Shrivastava, Suryakanth V Gangashetty and Anil Kumar Vuppala. We heartily express our gratitude to Sukomal Pal, Swasti Mishra and the great team of volunteers at IIT (BHU) and BHU, Varanasi for their timely help with sincere dedication and hard work to make this conference a success.

We also thank all those who came forward to help us in this task. We apologize if we have missed some names.

Finally, we thank all the researchers who responded to our call for papers and all the participants of ICON-2016, without whose overwhelming response the conference would not have been a success.

December 2016
Varanasi

Dipti Misra Sharma
Rajeev Sangal
Anil Kumar Singh

Advisory Committee:

Aravind K Joshi, University of Pennsylvania, USA (Chair)

Conference General Chair:

Rajeev Sangal, IIT (BHU), Varanasi, India

Programme Committee:

Anil Kumar Singh, IIT (BHU), India (Chair)
Dipti Misra Sharma, IIIT Hyderabad, India (Co-Chair)
Karunesh Arora, CDAC Noida, India
Sivaji Bandyopadhyay, Jadavpur University, Kolkata, India
Srinivas Bangalore, Interactions LLC, AT&T Research, USA
Peri Bhaskararao, IIIT Hyderabad, India
Rajesh Bhatt, University of Massachusetts, USA
Pushpak Bhattacharyya, IIT Patna, India
Ravindranath Chowdary C, IIT (BHU), India
Niladri Sekhar Dash, ISI Kolkata, India
Sanjukta Ghosh, BHU, Varanasi, India
Vishal Goyal, Punjabi University, Patiala, India
Harald Hammarstrom, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
Mohammed Hasanuzzaman, Université de Caen Normandie, Caen, France
Gerold Hintz, TU Darmstadt, Germany
Samar Husain, IIT Delhi, India
Amba Kulkarni, University of Hyderabad, India
Ramamoorthy L., CIIL, Mysore
Gurpreet Lehal, Punjabi University, Patiala, India
Roser Morante, VU University, Amsterdam, The Netherlands
Jose Moreno, Université de Caen Normandie, Caen, France
Joakim Nivre, Uppsala University, Sweden
Alexis Palmer, Heidelberg University, Germany
Martha Palmer, University of Colorado Boulder, USA
Soma Paul, IIIT Hyderabad, India
Jyoti Pawar, DCST, Goa University, India
Eugen Ruppert, TU Darmstadt, Germany
Sriparna Saha, IIT Patna, India
Shikhar Kr. Sarma, Gauhati University, India
Elizabeth Sherly, IIITM-K, Trivandrum, India
Sobha Lalitha Devi, AU-KBC, Chennai, India
Keh-Yih Su, Institute of Information Science, Academia Sinica, Taiwan

Anil Thakur, BHU, Varanasi, India
Vasudeva Varma, IIIT Hyderabad, India

Tools Contest Chairs:

POS Tagging for Code-Mixed Indian Social Media Text Rationale:

Amitav Das, IIIT, Sri City, India (Chair)

Student Paper Competition Chair:

Asif Ekbal, IIT-Patna, India

Organizing Committee:

Sukomal Pal, IIT (BHU), Varanasi, India (Chair)
Swasti Mishra, IIT (BHU), Varanasi, India

Organizers

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY

HYDERABAD

Indian Institute of Technology (BHU), Varanasi

NLPAI

Natural Language Processing Association of India

Linguistic Data Consortium for Indian Languages (LDC-IL)

Sponsors

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.

Abhijit Mishra
Aditya Joshi
Aishwarya N Reganti
Alexis Palmer
Alok Chakrabarty
Amba Kulkarni
Amitava Das
Amrith Krishna
Anil Kumar Singh
Anil Kumar Vuppala
Anil Thakur
Anoop Kunchukuttan
Anupam Jamatia
Aravind Ganapathiraju
Ashwini Vaidya
Asif Ekbal
Ayushi Dalmia
Ayushi Pandey
Balaji Jagan
Balamurali A R
Baskaran Sankaran
Bharat Ram Ambati
Bhuvana Narasimhan
Bibekananda Kundu
Bikash Gyawali
Bira Chandra Singh
Björn Gambäck
Brijesh Bhatt
C V Jawahar
Debasis Ganguly
Deepak Padmanabhan
Dhananjaya Gowda
Dimitris Mavroeidis
Dimple Paul
Dinesh Kumar Prabhakar
Dipankar Das
Dipasree Pal
Dipti Misra Sharma
Douwe Kiela
Dwaipayan Roy
Dwijen Rudrapal
Elizabeth Sherly
Erik Cambria
Eugen Ruppert
Ganesh Jawahar
Girish Palshikar
Gurpreet Singh Lehal
Harald Hammarström
Hema Murthy
Himanshu Sharma
Imran Rasheed
Irshad Bhat
Joakim Nivre
Joe Cheri
Jose Moreno
Jyoti Pareek
Jyoti Pawar
K V Ramakrishnamacharyulu
K V Subbarao
Kamal Garg
Karunesh Arora
Keh-Yih Su
Kishorjit Nongmeikapam
Kunal Chakma
Lars Bungum
Malhar Kulkarni
Manish Shrivastava
Marco Damonte
Martha Palmer
Mohammed Hasanuzzaman
Monojit Choudhury
Nayeemulla Khan A
Nikhil Pattisapu
Nikhilesh Bhatnagar
Niladri Sekhar Dash
Niraj Kumar
Owen Rambow
Paolo Rosso
Parth Gupta
Partha Talukdar
Pawan Goyal
Peri Bhaskararao
Pradeepika Verma
Pranaw Kumar
Prasha Shresta
Priya Radhakrishnan
Pruthwik Mishra
Pushpak Bhattacharyya
Radhika Mamidi
Raghava Krishnan
Rajendra Prasath
Rajendran S
Rajesh Bhatt
Rajiv Srivastava
Rakesh Balabantaray
Raksha Sharma
Ranjani Parthasarathi
Ratish Surendran
Ravindranath Chowdary C
Riyaz Bhat
Roser Morante
Rudramurthy V
S Arulmozi
Sachin Pawar
Sai Krishna Rallabandi
Sajini T
Samar Husain
Samudravijaya K
Sandipan Dandapat
Sandya Mannarswamy
Sanjukta Ghosh
Saptarshi Ghosh
Sara Tonelli
Satarupa Guha
Shashi Narayan
Shourya Roy
Shubhnandan Singh
Sivaji Bandyopadhyay
Sobha Lalitha Devi
Soma Paul
Srikanth Ronanki
Sriparna Saha
Sriram Chaudhury
Sruti Rallapalli
Subalalitha Navaneetha Krishnan
Sudip Kumar Naskar
Sukomal Pal
Sunayana Sitaram
Sutanu Chakraborti
Swapnil Chaudhari
T Pattabhi R K Rao
Taniya Mishra
Thamar Solorio
Thierry Declerck
Tushar Maheshwari
Umamaheswari E
Vasudeva Varma
Vasudevan N
Vijay Sunday Ram
Vinay Kumar Mittal
Vineet Chaitanya
Vishal Goyal
Vishnu G Sriram
Yuji Matsumo

Table of Contents

Keynote Lecture 1: Practical Use of Machine Translation in International Organizations
Bruno Pouliquen 1

Integrating WordNet for Multiple Sense Embeddings in Vector Semantics
David Foley and Jugal Kalita 2

Can SMT and RBMT Improve each other's Performance?- An Experiment with English-Hindi Translation
Debajyoty Banik, Sukanta Sen, Asif Ekbal and Pushpak Bhattacharyya 10

Composition of Compound Nouns Using Distributional Semantics
Kyra Yee and Jugal Kalita 20

Towards Building a SentiWordNet for Tamil
Abishek Kannan, Gaurav Mohanty and Radhika Mamidi 30

Extending AIDA framework by incorporating coreference resolution on detected mentions and pruning based on popularity of an entity
Samaikya Akarapu and C Ravindranath Chowdary 36

Sentence Based Discourse Classification for Hindi Story Text-to-Speech (TTS) System
Kumud Tripathi, Parakrant Sarkar and K. Sreenivasa Rao 46

Biomolecular Event Extraction using a Stacked Generalization based Classifier
Amit Majumder, Asif Ekbal and Sudip Kumar Naskar 55

Syntax and Pragmatics of Conversation: A Case of Bangla
Samir Karmakar and Soumya Sankar Ghosh 65

Dependency grammars as Haskell programs
Tomasz Obrebski 71

Improving Document Ranking using Query Expansion and Classification Techniques for Mixed Script Information Retrieval
Subham Kumar, Anwesh Sinha Ray, Sabyasachi Kamila, Asif Ekbal, Sriparna Saha and Pushpak Bhattacharyya 81

Feature based Sentiment Analysis using a Domain Ontology
Neha Yadav and C Ravindranath Chowdary 90

Cross-lingual transfer parser from Hindi to Bengali using delexicalization and chunking
Ayan Das, Agnivo Saha and Sudeshna Sarkar 99

Constraint Grammar-based conversion of Dependency Treebanks
Eckhard Bick 109

Meaning Matters: Senses of Words are More Informative than Words for Cross-domain Sentiment Analysis
Raksha Sharma, Sudha Bhingardive and Pushpak Bhattacharyya 115

POS Tagging Experts via Topic Modeling
Atreyee Mukherjee, Sandra Kübler and Matthias Scheutz 120

Graph theoretic interpretation of Bangla traditional grammar
Samir Karmakar, Sayantani Banerjee and Soumya Ghosh 129

A method for Automatic Text Summarization using Consensus of Multiple Similarity Measures and Ranking Techniques
Mukesh Kumar Jadon and Ayush Pareek 137

Automatic Translation of English Text to Indian Sign Language Synthetic Animations
Lalit Goyal and Vishal Goyal 144

Towards Deep Learning in Hindi NER: An approach to tackle the Labelled Data Sparsity
Vinayak Athavale, Shreenivas Bharadwaj, Monik Pamecha, Ameya Prabhu and Manish Shrivastava 154

Vaidya: A Spoken Dialog System for Health Domain
Prathyusha Danda, Brij Mohan Lal Srivastava and Manish Shrivastava 161

Cosmopolitan Mumbai, Orthodox Delhi, Techcity Bangalore:Understanding City Specific Societal Sentiment
Aishwarya N Reganti, Tushar Maheshwari, Upendra Kumar and Amitava Das 167

Keynote Lecture 2: Neural (and other Machine Learning) Approaches to Text Normalization
Richard Sproat 177

Wisdom of Students: A Consistent Automatic Short Answer Grading Technique
Shourya Roy, Sandipan Dandapat, Ajay Nagesh and Narahari Y. 178

A Recurrent Neural Network Architecture for De-identifying Clinical Records
Shweta, Ankit Kumar, Asif Ekbal, Sriparna Saha and Pushpak Bhattacharyya 188

Twitter Named Entity Extraction and Linking Using Differential Evolution
Utpal Kumar Sikdar and Björn Gambäck 198

Learning Non-Linear Functions for Text Classification
Cohan Sujay Carlos and Geetanjali Rakshit 208

A Computational Analysis of Mahabharata
Debarati Das, Bhaskarjyoti Das and Kavi Mahesh 219

Use of Features for Accentuation of ghañanta Words
Samir Janardan Sohoni and Malhar A. Kulkarni 229

Learning to Identify Subjective Sentences
Girish K. Palshikar, Manoj Apte, Deepak Pandita and Vikram Singh 239

Opinion Mining in a Code-Mixed Environment: A Case Study with Government Portals
Deepak Gupta, Ankit Lamba, Asif Ekbal and Pushpak Bhattacharyya 249

Use of Semantic Knowledge Base for Enhancement of Coherence of Code-mixed Topic-Based Aspect Clusters
Kavita Asnani and Jyoti D Pawar 259

Genetic Algorithm (GA) Implementation for Feature Selection in Manipuri POS Tagging
Kishorjit Nongmeikapam and Sivaji Bandyopadhyay 267

Effect of Syntactic Features in Bangla Sentence Comprehension
Manjira Sinha, Tirthankar Dasgupta and Anupam Basu 275

A New Feature Selection Technique Combined with ELM Feature Space for Text Classification
Rajendra Kumar Roul and Pranav Rai 285

On Why Coarse Class Classification is Bottleneck in Noun Compound Interpretation
Girishkumar Ponkiya, Pushpak Bhattacharyya and Girish K. Palshikar 293

Verbframator:Semi-Automatic Verb Frame Annotator Tool with Special Reference to Marathi
Hanumant Redkar, Sandhya Singh, Nandini Ghag, Jai Paranjape, Nilesh Joshi, Malhar Kulkarni and Pushpak Bhattacharyya 299

Towards Building A Domain Agnostic Natural Language Interface to Real-World Relational Databases
Sree Harsha Ramesh, Jayant Jain, Sarath K S and Krishna R Sundaresan 305

Experimental Study of Vowels in Nagamese, Ao and Lotha: Languages of Nagaland
Joyanta Basu, Tulika Basu, Soma Khan, Madhab Pal, Rajib Roy and Tapan Kumar Basu 315

Perception of Phi-Phrase boundaries in Hindi.
Somnath Roy 324

Conference Program

Monday, December 19, 2016

+ 9:00-9:30 Inaugural Ceremony

+ 9:30-10:00 Keynote Lecture 1 by Bruno Pouliquen

Keynote Lecture 1: Practical Use of Machine Translation in International Organizations
Bruno Pouliquen

+ 10:30-11:00 Tea Break

+ 11:00-13:00 Technical Session I: Machine Translation and WSD:

Integrating WordNet for Multiple Sense Embeddings in Vector Semantics
David Foley and Jugal Kalita

Can SMT and RBMT Improve each other's Performance?- An Experiment with English-Hindi Translation
Debajyoty Banik, Sukanta Sen, Asif Ekbal and Pushpak Bhattacharyya

Composition of Compound Nouns Using Distributional Semantics
Kyra Yee and Jugal Kalita

Towards Building a SentiWordNet for Tamil
Abishek Kannan, Gaurav Mohanty and Radhika Mamidi

+ 11:00-13:00 Technical Session II : Discourse Analysis:

Extending AIDA framework by incorporating coreference resolution on detected mentions and pruning based on popularity of an entity
Samaikya Akarapu and C Ravindranath Chowdary

Sentence Based Discourse Classification for Hindi Story Text-to-Speech (TTS) System
Kumud Tripathi, Parakrant Sarkar and K. Sreenivasa Rao

Biomolecular Event Extraction using a Stacked Generalization based Classifier
Amit Majumder, Asif Ekbal and Sudip Kumar Naskar

Monday, December 19, 2016 (continued)

Syntax and Pragmatics of Conversation: A Case of Bangla
Samir Karmakar and Soumya Sankar Ghosh

+ 13:00-14:00 Lunch

+ 14:00-15:00 Technical Session III: New Trends:

Dependency grammars as Haskell programs
Tomasz Obrebski

Improving Document Ranking using Query Expansion and Classification Techniques for Mixed Script Information Retrieval
Subham Kumar, Anwesh Sinha Ray, Sabyasachi Kamila, Asif Ekbal, Sriparna Saha and Pushpak Bhattacharyya

+ 15:00-15:30 Tea Break

+ 15:30-17:30 Poster and Demo Session-1:

Feature based Sentiment Analysis using a Domain Ontology
Neha Yadav and C Ravindranath Chowdary

Cross-lingual transfer parser from Hindi to Bengali using delexicalization and chunking
Ayan Das, Agnivo Saha and Sudeshna Sarkar

Constraint Grammar-based conversion of Dependency Treebanks
Eckhard Bick

Meaning Matters: Senses of Words are More Informative than Words for Cross-domain Sentiment Analysis
Raksha Sharma, Sudha Bhingardive and Pushpak Bhattacharyya

POS Tagging Experts via Topic Modeling
Atreyee Mukherjee, Sandra Kübler and Matthias Scheutz

Graph theoretic interpretation of Bangla traditional grammar
Samir Karmakar, Sayantani Banerjee and Soumya Ghosh

A method for Automatic Text Summarization using Consensus of Multiple Similarity Measures and Ranking Techniques
Mukesh Kumar Jadon and Ayush Pareek

Monday, December 19, 2016 (continued)

Automatic Translation of English Text to Indian Sign Language Synthetic Animations
Lalit Goyal and Vishal Goyal

Towards Deep Learning in Hindi NER: An approach to tackle the Labelled Data Sparsity
Vinayak Athavale, Shreenivas Bharadwaj, Monik Pamecha, Ameya Prabhu and Manish Shrivastava

Vaidya: A Spoken Dialog System for Health Domain
Prathyusha Danda, Brij Mohan Lal Srivastava and Manish Shrivastava

Cosmopolitan Mumbai, Orthodox Delhi, Techcity Bangalore:Understanding City Specific Societal Sentiment
Aishwarya N Reganti, Tushar Maheshwari, Upendra Kumar and Amitava Das

+ 17:30-18:30 NLPAI Meeting

+ 18:30-19:30 Cultural Program

+ 19:30-20:30 Dinner

Tuesday, December 20, 2016

+ 9:30-10:30 Keynote Lecture 2:

Keynote Lecture 2: Neural (and other Machine Learning) Approaches to Text Normalization
Richard Sproat

+ 10:30-11:00 Tea Break

Tuesday, December 20, 2016 (continued)

+ 11:00-13:00 Technical Session IV: Statistical Methods:

Wisdom of Students: A Consistent Automatic Short Answer Grading Technique
Shourya Roy, Sandipan Dandapat, Ajay Nagesh and Narahari Y.

A Recurrent Neural Network Architecture for De-identifying Clinical Records
Shweta, Ankit Kumar, Asif Ekbal, Sriparna Saha and Pushpak Bhattacharyya

Twitter Named Entity Extraction and Linking Using Differential Evolution
Utpal Kumar Sikdar and Björn Gambäck

Learning Non-Linear Functions for Text Classification
Cohan Sujay Carlos and Geetanjali Rakshit

+ 13:00-14:00 Lunch

+ 14:-15:30 Poster and Demo Session-2:

A Computational Analysis of Mahabharata
Debarati Das, Bhaskarjyoti Das and Kavi Mahesh

Use of Features for Accentuation of ghañanta Words
Samir Janardan Sohoni and Malhar A. Kulkarni

Learning to Identify Subjective Sentences
Girish K. Palshikar, Manoj Apte, Deepak Pandita and Vikram Singh

Opinion Mining in a Code-Mixed Environment: A Case Study with Government Portals
Deepak Gupta, Ankit Lamba, Asif Ekbal and Pushpak Bhattacharyya

Use of Semantic Knowledge Base for Enhancement of Coherence of Code-mixed Topic-Based Aspect Clusters
Kavita Asnani and Jyoti D Pawar

Genetic Algorithm (GA) Implementation for Feature Selection in Manipuri POS Tagging
Kishorjit Nongmeikapam and Sivaji Bandyopadhyay

Tuesday, December 20, 2016 (continued)

Effect of Syntactic Features in Bangla Sentence Comprehension
Manjira Sinha, Tirthankar Dasgupta and Anupam Basu

A New Feature Selection Technique Combined with ELM Feature Space for Text Classification
Rajendra Kumar Roul and Pranav Rai

On Why Coarse Class Classification is Bottleneck in Noun Compound Interpretation
Girishkumar Ponkiya, Pushpak Bhattacharyya and Girish K. Palshikar

Verbframator:Semi-Automatic Verb Frame Annotator Tool with Special Reference to Marathi
Hanumant Redkar, Sandhya Singh, Nandini Ghag, Jai Paranjape, Nilesh Joshi, Malhar Kulkarni and Pushpak Bhattacharyya

Towards Building A Domain Agnostic Natural Language Interface to Real-World Relational Databases
Sree Harsha Ramesh, Jayant Jain, Sarath K S and Krishna R Sundaresan

+ 15:30-16:00 Tea Break

+ 16:00-17:30 Technical Session V: Speech Processing:

Experimental Study of Vowels in Nagamese, Ao and Lotha: Languages of Nagaland
Joyanta Basu, Tulika Basu, Soma Khan, Madhab Pal, Rajib Roy and Tapan Kumar Basu

Perception of Phi-Phrase boundaries in Hindi.
Somnath Roy

+ 16:00-17:30 Technical Session VI : NLP Tool Contest

Tuesday, December 20, 2016 (continued)

+ 17:30-18:00 Valedictory Session