ISCLS 2019

Proceedings of the 6th International Sanskrit Computational Linguistics Symposium

23–25 October, 2019
Indian Institute of Technology Kharagpur
West Bengal, India

Sponsors

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-950737-71-0

Introduction

Welcome to the 6th edition of the International Sanskrit Computational Linguistics Symposium (6th ISCLS) at IIT Kharagpur, West Bengal, India. The aim of ISCLS is to bring together researchers interested in any aspects of Sanskrit Computational Linguistics. Full papers were invited on original and unpublished research on various aspects of Computational Linguistics and Digital Humanities related to Sanskrit (Classical and Vedic), Prakrit, Pali, Buddhist Hybrid Sanskrit, etc. 13 contributions were accepted, and the final versions, after incorporating the reviewers' comments constitute the proceedings. We would like to thank the Program Committee for the 6th ISCLS for their reviewing efforts:

- Stefan Baums (University of Munich)
- Laxmidhar Behera (IIT Kanpur)
- Brendan Gillon (McGill University)
- Pawan Goyal (IIT Kharagpur)
- Olivier Hellwig (University of Zurich)
- Gérard Huet (Inria Paris)
- Amba Kulkarni (University of Hyderabad)
- Malhar Kulkarni (IIT Bombay)
- Pavan Kumar Satuluri (Chinmaya Vishwavidyapeeth, Veliyanad)
- Andrew Ollett (Harvard University, Cambridge, MA)
- Dhaval Patel (Ahmedabad)
- Ganesh Ramakrishnan (IIT Bombay)
- Peter Scharf (IIIT Hyderabad)
- Srinivasa Varakhedi (KKSU, Ramtek)
- K Varalakshmi (Osmania University, Hyderabad)

The first two papers talk about Sanskrit sentence generation and parsing. In "Sanskrit Sentence Generator", Amba Kulkarni and Madhusoodana Pai J present a sentence generator for Sanskrit, which takes an intermediate representation from which, using Panini's grammar, the desired sentence can be generated, without appealing to the world knowledge. In "Dependency Parser for Sanskrit Verses", Amba Kukarni, Sanal Vikram and Sriram K describe their efforts to build a dependency parser which parses both prose as well as verse texts. The parser utilizes various constraints following traditional rules of verbal cognition, which are employed using and edge-centric binary join method.

The next two papers discuss the compound identification and type classification using word embeddings and machine learning methods. The paper, "Revisiting the Role of Feature Engineering for Compound Type Identification in Sanskrit" by Jivnesh Sandhan, Amrith Krishna, Pawan Goyal and Laxmidhar Behera, attempts to ask the question if the recent advances in neural networks can outperform traditional

hand engineered feature based methods on the semantic level multi-class classification task for Sanskrit. In “A Machine Learning Approach for Identifying Compound Words from a Sanskrit Text", Premjith B, Chandni Chandran V, Shriganesh Bhat, Soman Kp and Prabaharan P propose a classification framework for finding the compound words from a Sanskrit text, in particular, those found in Ayurveda text books, using word embeddings.

The next two papers talk about NLP corpus building. In “LDA Topic Modeling for pramāṅa Texts: A Case Study in Sanskrit NLP Corpus Building", Tyler Neill describes the methodology followed towards the preparation of digital corpus for word-level analysis. It also explains pitfalls in current digitalization practices of Sanskrit corpus. In “Vedavaapi: A Platform for Community-sourced Indic Knowledge Processing at Scale", Sai Susarla and Damodar Reddy Challa describe the architecture of an online platform for end-to-end indic knowledge processing addressing the challenges of composing independently developed tools for higher-level tasks, as well as employing human experts in the loop to work around the limitations of automated tools.

The next two contributions discuss the problems concerning information retrieval and questions answering from Sanskrit texts. The paper, “On Sanskrit and Information Retrieval" by Michaël Meyer discusses the challenges for traditional information retrieval systems to handle the peculiarities of Sanskrit, and discusses a few possible solutions. In “Framework for Question-Answering in Sanskrit through Automated Construction of Knowledge Graphs", Hrishikesh Terdalkar and Arnab Bhattacharya target the problem of building knolwedge graphs for particular types of relations from Sanskrit texts and attempts to answer factoid questions using the extracted relations.

The next two papers discuss digital tools for Sanskrit Wordnet and Vaijayantīkośa. In “Introduction to Sanskrit Shabdamitra: An Educational Application of Sanskrit Wordnet", Malhar Kulkarni, Nilesh Joshi, Sayali Khare, Hanumant Redkar and Pushpak Bhattacharyya introduce Sanskrit Shabdamitra, a digital tool based on Sanskrit Wordnet, for learning and teaching Sanskrit. The paper, “Vaijayantīkośa Knowledge-Net" by Aruna Vayuvegula, Satish Kanugovi, Sivaja S Nair, Shivani V and Mahalakshmi discusses Vaijayantīkośa Knowledge-Net, a web-based tool for easy access and analysis of words in Vaijayantīkośa, a Sanskrit lexicon containing words from spoken language as well as those in Vedic literature.

The next two contributions attempt to capture the evolution of manuscript texts. The paper, “Utilizing Word Embeddings based Features for Phylogenetic Tree Generation of Sanskrit Texts" by Diptesh Kanojia, Abhijeet Dubey, Malhar Kulkarni, Pushpak Bhattacharyya and Reza Haffari infers phylogenetic trees of Sanskrit texts using inter-manuscript distances obtained via word embeddings. In “An Introduction to the Textual History Tool", Diptesh Kanojia, Malhar Kulkarni, Pushpak Bhattacharyya, Sayali Ghodekar, Irawati Kulkarni, Nilesh Joshi and Eivind Kahrs describe textual history tool to capture the historical view of the transmission of a text through the manuscript tradition, captured via inter-related data from various types of related texts.

The proceedings conclude with the paper, “Pāli Sandhi – A Computational Approach" by Swati Basapur, Shivani V and Sivaja Nair, which discusses complexities involved in creating a computational grammar for Sandhi tools in Pāli language.

ISCLS 2019 has received financial support from Dharohar, Indic-Academy and DST-SERB.

The conference also hosts two keynote talks by Prof. Rajeev Sangal and Prof. Korada Subrahmanyam. Further, various demo submissions are also presented at the conference.

We very much hope that you will have an enjoyable and inspiring time at the conference!

Pawan Goyal
Indian Institute of Technology, Kharagpur, WB, India
October 2019

Table of Contents