LREC-COLING 2024

The 7th Workshop on Indian Language Data Resource and Evaluation @LREC-COLING-2024 (WILDRE-7)

Workshop Proceedings

Editors
Girish Jha, Sobha Lalitha Devi, Kalika Bali and Atul Kr. Ojha

25 May, 2024
Torino, Italia

Proceedings of the 7th Workshop on Indian Language Data Resource and Evaluation @LREC-COLING-2024 (WILDRE-7)

ISBN 978-2-493814-37-1
ISSN 2951-2093 (COLING); 2522-2686 (LREC)

Jointly organized by the ELRA Language Resources Association
and the International Committee on Computational Linguistics

Preface

WILDRE – the 7th Workshop on Indian Language Data: Resources and Evaluation is being organized in Torino, Italia on May 25th, 2024 under the LREC-COLING 2024 platform. India has a huge linguistic diversity and has seen concerted efforts from the Indian government and industry towards developing language resources. ELRA Language Resources Association and its associate organizations have been very active and successful in addressing the challenges and opportunities related to language resource creation and evaluation. It is, therefore, a great opportunity for resource creators of Indian languages to showcase their work on this platform and also to interact and learn from those involved in similar initiatives all over the world.

The broader objectives of the 7th WILDRE will be

- to map the status of Indian Language Resources
- to investigate challenges related to creating and sharing various levels of language resources
- to promote a dialogue between language resource developers and users
- to provide an opportunity for researchers from India to collaborate with researchers from other parts of the world

The call for papers received a good response from the Indian language technology community. In addition, WILDRE-7 included a Shared Task on Code-mixed Less-resourced Sentiment Analysis for Indo-Aryan Languages.

This year, we selected only three papers for oral, one findings paper and eight for poster presentations (including two system descriptions and one non-archival).

Workshop Organisers

Workshop Chairs

Girish Nath Jha, Chairman, Commission for Scientific and Technical Terminology, MoE, GOI and JNU, New Delhi
Kalika Bali, Microsoft Research India Lab, Bangalore
Sobha L, AU-KBC, Anna University
Atul Kr. Ojha, University of Ireland Galway, Ireland

Program Committee

Anil Kumar Singh, IIT BHU, Benaras
Anoop Kunchukuttan, Microsoft AI and Research, India
Anupam Basu, Director, NIIT, Durgapur
Arulmozi Selvaraj, University of Hyderabad
Asif Iqbal, IIT Patna, Patna
Atul Kr. Ojha, University of Ireland Galway, Ireland & Panlingua Language Processing LLP, India
Bogdan Babych, Heidelberg University, Germany
Daan van Esch, Google, USA
Dafydd Gibbon, Universität Bielefeld, Germany
Dipti Mishra Sharma, IIIT-Hyderabad
Elizabeth Sherley, IITM-Kerala, Trivandrum
Gaurav Negi, University of Galway
Georg Rehm, DFKI, Germany
Girish Nath Jha, Chairman, Commission for Scientific and Technical Terminology, MoE, GOI and JNU, New Delhi
Jolanta Bachan, Adam Mickiewicz University, Poland
Joseph Mariani, LIMSI-CNRS, France
Khalid Choukri, ELRA, France
Lars Hellan, NTNU, Norway
Manji Bhadra, Bankura University, West Bengal
Malhar Kulkarni, IIT Bombay
Massimo Moneglia, University of Florence, Italy
Monojit Choudhary, Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi
Narayan Choudhary, CIIL, Mysore
Niladri Shekhar Dash, ISI Kolkata
Panchanan Mohanty, GLA, Mathura
Priya Rani, University of Galway
Rajeev R R, ICFOSS, Trivandrum
Shantipriya Parida, Silo AI, Finland
Shagun Sinha, Amity University, Noida, India
Shivaji Bandhopadhyay, Jadavpur University
Sobha Lalitha Devi, AU-KBC Research Centre, Anna University
Subhash Chandra, Delhi University
Swaran Lata, Retired Head, TDIL, MCIT, Govt of India
Virach Sornlertlamvanich, Thammasat Univeristy, Bangkok, Thailand
Zygmunt Vetulani, Adam Mickiewicz University, Poland

Table of Contents

Conference Program

Saturday, May 25, 2024

14:00–14:05 ***Welcome by Workshop Chairs***

14:05–15:00 *Keynote Lecture*
TBD

15:00–16:00 **Oral Session-I**

15:00–15:25 *Towards Disfluency Annotated Corpora for Indian Languages*
Chayan Kochar, Vandan Vasantlal Mujadia, Pruthwik Mishra and Dipti Misra Sharma

15:25–15:45 *EmoMix-3L: A Code-Mixed Dataset for Bangla-English-Hindi for Emotion Detection*
Nishat Raihan, Dhiman Goswami, Antara Mahmud, Antonios Anastasopoulos and Marcos Zampieri

15:45–16:00 *Findings of the WILDRE Shared Task on Code-mixed Less-resourced Sentiment Analysis for Indo-Aryan Languages*
Priya Rani, Gaurav Negi, Saroj Jha, Shardul Suryawanshi, Atul Kr. Ojha, Paul Buitelaar and John P. McCrae

16:00–16:30 **Coffee break/Poster Session**

16:00–16:30 *Multilingual Bias Detection and Mitigation for Indian Languages*
Ankita Maity, Anubhav Sharma, Rudra Dhar, Tushar Abhishek, Manish Gupta and Vasudeva Varma

16:00–16:30 *Dharmaśāstra Informatics: Concept Mining System for Socio-Cultural Facet in Ancient India*
Arooshi Nigam and Subhash Chandra

16:00–16:30 *Exploring News Summarization and Enrichment in a Highly Resource-Scarce Indian Language: A Case Study of Mizo*
Abhinaba Bala, Ashok Urlana, Rahul Mishra and Parameswari Krishnamurthy

16:00–16:30 *Finding the Causality of an Event in News Articles*
Sobha Lalitha Devi and Pattabhi RK Rao

16:00–16:30 *Creating Corpus of Low Resource Indian Languages for Natural Language Processing: Challenges and Opportunities*
Pratibha Dongare

Saturday, May 25, 2024 (continued)

16:00–16:30 *FZZG at WILDRE-7: Fine-tuning Pre-trained Models for Code-mixed, Less-resourced Sentiment Analysis*
Gaurish Thakkar, Marko Tadić and Nives Mikelic Preradovic

16:00–16:30 *MLInitiative@WILDRE7: Hybrid Approaches with Large Language Models for Enhanced Sentiment Analysis in Code-Switched and Code-Mixed Texts*
Hariram Veeramani, Surendrabikram Thapa and Usman Naseem

16:30–16:55 Oral Session-II

16:30–16:55 *Aalamaram: A Large-Scale Linguistically Annotated Treebank for the Tamil Language*
A M Abirami, Wei Qi Leong, Hamsawardhini Rengarajan, D Anitha, R Suganya, Himanshu Singh, Kengatharaiyer Sarveswaran, William Chandra Tjhi and Rajiv Ratn Shah

16:55–17:40 Panel discussion

17:40–17:50 ***Valedictory Sessioner a title here***

17:50–17:55 ***Vote of Thanks***