

6th Workshop on Indian Language Data: Resources and Evaluation (WILDRE-6)

Held at LREC 2022

Marseille, France
20 – 25 June 2022

Editors:

**Girish Nath Jha
Sobha L**

**Kalika Bali
Atul Kr. Ojha**

ISBN: 978-1-7138-6143-0

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2022) by European Language Resources Association (ELRA)
All rights reserved.

Copyright for individual papers remains with the authors and are licensed under a Creative Commons 4.0 license, CC-BY-NC. (<https://creativecommons.org/licenses/by-nc/4.0/>)

Printed with permission by Curran Associates, Inc. (2023)

For permission requests, please contact the Association for Computational Linguistics at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>Introducing EM-FT for Manipuri-English Neural Machine Translation</i> Rudali Huidrom and Yves Lepage	1
<i>L3Cube-HingCorpus and HingBERT: A Code Mixed Hindi-English Dataset and BERT Language Models</i> Ravindra Nayak and Raviraj Joshi	7
<i>Leveraging Sub Label Dependencies in Code Mixed Indian Languages for Part-Of-Speech Tagging using Conditional Random Fields.</i> Akash Kumar Gautam	13
<i>HindiWSD: A package for word sense disambiguation in Hinglish & Hindi</i> Mirza Yusuf, Praatibh Surana and Chethan sharma	18
<i>Pāṇinian Phonological Changes: Computation and Development of Online Access System</i> Sanju . and Subhash Chandra	24
<i>L3Cube-MahaNER: A Marathi Named Entity Recognition Dataset and BERT models</i> Onkar Litake, Maithili Ravindra Sabane, Parth Sachin Patil, Aparna Abhijeet Ranade and Raviraj Joshi	29
<i>Identifying Emotions in Code Mixed Hindi-English Tweets</i> Sanket Sonu, Rejwanul Haque, Mohammed Hasanuzzaman, Paul Stynes and Pramod Pathak . . .	35
<i>Digital Accessibility and Information Mining of Dharmasāstric Knowledge Traditions</i> Arooshi Nigam and Subhash Chandra	42
<i>Language Resource Building and English-to-Mizo Neural Machine Translation Encountering Tonal Words</i> Vanlalmuansangi Khenglawt, Sahinur Rahman Laskar, Santanu Pal, Partha Pakray and Ajoy Kumar Khan	48
<i>Classification of Multiword Expressions in Malayalam</i> Treesa Cyriac and Sobha Lalitha Devi	55
<i>Bengali and Magahi PUD Treebank and Parser</i> Pritha Majumdar, Deepak Alok, Akanksha Bansal, Atul Kr. Ojha and John P. McCrae	60
<i>Makadi: A Large-Scale Human-Labeled Dataset for Hindi Semantic Parsing</i> Shashwat Vaibhav and Nisheeth Srivastava	68
<i>Automatic Identification of Explicit Connectives in Malayalam</i> Kumari Sheeja S and Sobha Lalitha Devi	74
<i>Web based System for Derivational Process of Kṛdanta based on Pāṇinian Grammatical Tradition</i> Sumit Sharma and Subhash Chandra	80
<i>Universal Dependency Treebank for Odia Language</i> Shantipriya Parida, Kalyanamalini Shabadi, Atul Kr. Ojha, Saraswati Sahoo, Satya Ranjan Dash and Bijayalaxmi Dash	84
<i>Computational Referencing System for Sanskrit Grammar</i> Baldev Khandoliyan and Ram Kishor	90

L3Cube-MahaCorpus and MahaBERT: Marathi Monolingual Corpus, Marathi BERT Language Models, and Resources
Raviraj Joshi 97