5th Workshop on Indian Language Data: Resources and Evaluation (WILDRE5)

Marseille, France
11-16 May 2020

Editors:

Girish Nath Jha
Kalika Bali
Sobha L
S. S. Agrawal
Atul Kr. Oj

ISBN: 978-1-7138-1280-7

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© (2020) by the Association for Computational Linguistics
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. (2020)

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

Part-of-Speech Annotation Challenges in Marathi
Gajanan Rane, Nilesh Joshi, Geetanjali Rane, Hanumant Redkar, Malhar Kulkarni and Pushpak Bhattacharyya 1

A Dataset for Troll Classification of TamilMemes
Shardul Suryawanshi, Bharathi Raja Chakravarthi, Pranav Verma, Mihael Arcan, John Philip McCrae and Paul Buitelaar 7

OdiEnCorp 2.0: Odia-English Parallel Corpus for Machine Translation
Shantipriya Parida, Satya Ranjan Dash, Ondřej Bojar, Petr Motlicek, Priyanka Pattnaik and Debasish Kumar Mallick 14

Handling Noun-Noun Coreference in Tamil
Vijay Sundar Ram and Sobha Lalitha Devi 20

Malayalam Speech Corpus: Design and Development for Dravidian Language
Lekshmi K R, Jithesh V S and Elizabeth Sherly 25

Multilingual Neural Machine Translation involving Indian Languages
Pulkit Madaan and Fatiha Sadat 29

Universal Dependency Treebanks for Low-Resource Indian Languages: The Case of Bhojpuri
Atul Kr. Ojha and Daniel Zeman 33

A Fully Expanded Dependency Treebank for Telugu
Sneha Nallani, Manish Shrivastava and Dipti Sharma 39

Determination of Idiomatic Sentences in Paragraphs Using Statement Classification and Generalization of Grammar Rules
Naziya Shaikh 45

Polish Lexicon-Grammar Development Methodology as an Example for Application to other Languages
Zygmunt Vetulani and Grażyna Vetulani 51

Abstractive Text Summarization for Sanskrit Prose: A Study of Methods and Approaches
Shagun Sinha and Girish Jha 60

A Deeper Study on Features for Named Entity Recognition
Malarkodi C S and Sobha Lalitha Devi 66