

First Workshop on South East Asian Language Processing 2023

Nusa Dua, Indonesia
1 November 2023

ISBN: 978-1-7138-9008-9

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© (2023) by the Asian Federation of Natural Language Processing and the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2024)

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>Towards Automatic Construction of Filipino WordNet: Word Sense Induction and Synset Induction Using Sentence Embeddings</i>	
Dan John Velasco, Axel Alba, Trisha Gail Pelagio, Bryce Anthony Ramirez, Jan Christian Blaise Cruz, Unisse Chua, Briane Paul Samson and Charibeth Cheng	1
<i>Developing a Named Entity Recognition Dataset for Tagalog</i>	
Lester James Miranda	13
<i>Balarila: Deep Learning for Semantic Grammar Error Correction in Low-Resource Settings</i>	
Andre Dominic H. Ponce, Joshue Salvador A. Jadie, Paolo Edni Andryn Espiritu and Charibeth Cheng	21
<i>Utilizing Weak Supervision to Generate Indonesian Conversation Datasets</i>	
Mega Fransiska, Diah Pitaloka, Saripudin Saripudin, Satrio Putra and Lintang Sutawika*	30
<i>InstructAlign: High-and-Low Resource Language Alignment via Continual Crosslingual Instruction Tuning</i>	
Samuel Cahyawijaya, Holy Lovenia, Tiezheng Yu, Willy Chung and Pascale Fung	55
<i>SentMix-3L: A Novel Code-Mixed Test Dataset in Bangla-English-Hindi for Sentiment Analysis</i>	
Md Nishat Raihan, Dhiman Goswami, Antara Mahmud, Antonios Anastasopoulos and Marcos Zampieri	79
<i>IndoToD: A Multi-Domain Indonesian Benchmark For End-to-End Task-Oriented Dialogue Systems</i>	
Muhammad Kautsar, Rahmah Nurdini, Samuel Cahyawijaya, Genta Winata and Ayu Purwarianti	85
<i>Replicable Benchmarking of Neural Machine Translation (NMT) on Low-Resource Local Languages in Indonesia</i>	
Lucky Susanto, Ryandito Diandaru, Adila Krisnadhi, Ayu Purwarianti and Derry Tanti Wijaya	100