

6th Workshop on South and Southeast Asian Natural Language Processing (WSSANLP 2016)

Osaka, Japan
11 – 16 December 2016

ISBN: 978-1-5108-3317-3

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© (2016) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2017)

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

Full Papers

<i>Compound Type Identification in Sanskrit: What Roles do the Corpus and Grammar Play?</i> Amrith Krishna, Pavankumar Satuluri, Shubham Sharma, Apurv Kumar and Pawan Goyal	1
<i>Comparison of Grapheme-to-Phoneme Conversion Methods on a Myanmar Pronunciation Dictionary</i> Ye Kyaw Thu, Win Pa Pa, Yoshinori Sagisaka and Naoto Iwahashi	11
<i>Character-Aware Neural Networks for Arabic Named Entity Recognition for Social Media</i> Mourad Gridach	23
<i>Development of a Bengali parser by cross-lingual transfer from Hindi</i> Ayan Das, Agnivo Saha and Sudeshna Sarkar	33
<i>Sinhala Short Sentence Similarity Calculation using Corpus-Based and Knowledge-Based Similarity Measures</i> Jes Kadupitiya, Surangika Ranathunga and Gihan Dias	44
<i>Enriching Source for English-to-Urdu Machine Translation</i> Bushra Jawaid, Amir Kamran and Ondřej Bojar	54
<i>The IMAGACT4ALL Ontology of Animated Images: Implications for Theoretical and Machine Translation of Action Verbs from English-Indian Languages</i> Pitambar Behera, Sharmin Muzaffar, Atul kr. Ojha and Girish Jha	64
<i>Crowdsourcing-based Annotation of Emotions in Filipino and English Tweets</i> Fermin Roberto Lapitan, Riza Theresa Batista-Navarro and Eliezer Albacea	74
<i>Temporal Information Extraction in Clinical Domain (TIECA)</i> Shujeevan Kanapathipillai, Viraj Welgama and Ruwan Weerasinghe	83
<i>Sentiment Analysis of Tweets in Three Indian Languages</i> Shanta Phani, Shibamouli Lahiri and Arindam Biswas	93
<i>Dealing with Linguistic Divergences in English-Bhojpuri Machine Translation</i> Pitambar Behera, Neha Mourya and Vandana Pandey	103
<i>The development of a web corpus of Hindi language and corpus-based comparative studies to Japanese</i> Miki Nishioka and Shiro Akasegawa	114
<i>Automatic Creation of a Sentence Aligned Sinhala-Tamil Parallel Corpus</i> Riyafa Abdul Hameed, Nadeeshani Pathirennehelage, Anusha Ihalapathirana, Maryam Ziyad Mohamed, Surangika Ranathunga, Sanath Jayasena, Gihan Dias and Sandareka Fernando	124
<i>Clustering-based Phonetic Projection in Mismatched Crowdsourcing Channels for Low-resourced ASR</i> Wenda Chen, Mark Hasegawa-Johnson, Nancy Chen, Preethi Jyothi and Lav Varshney	133
<i>Improving the Morphological Analysis of Classical Sanskrit</i> Oliver Hellwig	142

<i>Query Translation for Cross-Language Information Retrieval using Multilingual Word Clusters</i> Paheli Bhattacharya, Pawan Goyal and Sudeshna Sarkar	152
<i>A study of attention-based neural machine translation model on Indian languages</i> Ayan Das, Pranay Yerra, Ken Kumar and Sudeshna Sarkar	163
<i>Comprehensive Part-Of-Speech Tag Set and SVM based POS Tagger for Sinhala</i> Sandareka Fernando, Surangika Ranathunga, Sanath Jayasena and Gihan Dias	173

Short Papers

<i>Align Me: A framework to generate Parallel Corpus Using OCRs and Bilingual Dictionaries</i> Priyam Bakliwal, Devadath V V and C V Jawahar	183
<i>Learning Indonesian-Chinese Lexicon with Bilingual Word Embedding Models and Monolingual Signals</i> Xinying Qiu and Gangqin Zhu	188
<i>Creating rich online dictionaries for the Lao–French language pair, reusable for Machine Translation</i> Vincent Berment	194