

# Communications in Computer and Information Science

781

*Commenced Publication in 2007*

Founding and Former Series Editors:

Alfredo Cuzzocrea, Xiaoyong Du, Orhun Kara, Ting Liu, Dominik Ślęzak,  
and Xiaokang Yang

## Editorial Board

Simone Diniz Junqueira Barbosa

*Pontifical Catholic University of Rio de Janeiro (PUC-Rio),  
Rio de Janeiro, Brazil*

Phoebe Chen

*La Trobe University, Melbourne, Australia*

Joaquim Filipe

*Polytechnic Institute of Setúbal, Setúbal, Portugal*

Igor Kotenko

*St. Petersburg Institute for Informatics and Automation of the Russian  
Academy of Sciences, St. Petersburg, Russia*

Krishna M. Sivalingam

*Indian Institute of Technology Madras, Chennai, India*

Takashi Washio

*Osaka University, Osaka, Japan*

Junsong Yuan

*Nanyang Technological University, Singapore, Singapore*

Lizhu Zhou

*Tsinghua University, Beijing, China*

More information about this series at <http://www.springer.com/series/7899>


Kôiti Hasida · Win Pa Pa (Eds.)

# Computational Linguistics

15th International Conference of the Pacific Association  
for Computational Linguistics, PACLING 2017  
Yangon, Myanmar, August 16–18, 2017  
Revised Selected Papers

*Editors*

Kôiti Hasida  
Graduate School of Information Science  
and Technology  
The University of Tokyo  
Tokyo  
Japan

Win Pa Pa   
Natural Language Processing Lab  
University of Computer Studies, Yangon  
Yangon  
Myanmar

ISSN 1865-0929 ISSN 1865-0937 (electronic)  
Communications in Computer and Information Science  
ISBN 978-981-10-8437-9 ISBN 978-981-10-8438-6 (eBook)  
<https://doi.org/10.1007/978-981-10-8438-6>

Library of Congress Control Number: 2018935886

© Springer Nature Singapore Pte Ltd. 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.  
part of Springer Nature  
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721,  
Singapore

## Preface

This volume is a compilation of selected papers from PACLING 2017, which was the 15th in the series of conferences held since 1989. Most of these events were held in Australia, Japan, and Canada, but recently we have gathered in developing countries: Malaysia in 2011, Indonesia in 2015, and Myanmar this time.

This shift coincides with the prospect that these latter countries, among others, have bigger near-future potentials in realizing smart societies powered by the circulation of rich and abundant data. For instance, India is building a presence-less, paperless, and cashless service infrastructure consisting of a nationwide person authentication system, open-API private and public services, a national PDS (personal data store) for each individual to coordinate these services while utilizing their personal data. Cambodia also aims at a cashless society based on a still lighter infrastructure.

It is very likely that most other Asian countries share similar visions of near-future societies powered by data circulation. Such restructurings are not only inexpensive enough for developing countries, but also much easier in those countries than in advanced countries facing stronger opposition by many vested interests. I hence feel this year's PACLING in Myanmar, with a much higher literacy rate than in India and Cambodia, is exciting.

I hope the conference and the proceedings contribute to the construction of smart societies, as technologies to deal with language could address essential parts of the data circulation.

December 2017

Kôiti Hasida

# Organization

## Honorary Chair

Kôiti Hasida                      The University of Tokyo, Japan

## Local Organizing Chair

Mie Mie Thet Thwin              University of Computer Studies, Yangon, Myanmar

## Local Organizing Committee

Nang Saing Moon Kham          University of Computer Studies, Yangon, Myanmar  
Win Pa Pa                          University of Computer Studies, Yangon, Myanmar

## Program Committee

Kenji Araki                          Hokkaido University, Japan  
Eiji Aramaki                        NAIST, Japan  
Normaziah Abdul Aziz          International Islamic University, Malaysia  
Tetsuro Chino                      Toshiba Corporation, Japan  
Khalid Choukri                     ELRA/ELDA, France  
Koji Dosaka                         Akita Prefectural University, Japan  
Alexander Gelbukh                Instituto Politécnico Nacional (IPN), Mexico  
Li Haizhou                          National University of Singapore, Singapore  
Yoshihiko Hayashi                Waseda University, Japan  
Tin Myat Htwe                      KyaingTon Computer University, Myanmar  
Bowen Hui                         University of British Columbia, Canada  
Kentaro Inui                        Tohoku University, Japan  
Kai Ishikawa                        NEC Corporation, Japan  
Hiroyuki Kameda                 Tokyo University of Technology, Japan  
Vlado Kesel                        Dalhousie University, Canada  
Satoshi Kinoshita                 JAPIO, Japan  
Kiyoshi Kogure                    Kanazawa Institute of Technology, Japan  
Qin Lu                                The Hong Kong Polytechnic University, SAR China  
Joseph Mariani                     LIMSI-CNRS, France  
Robert Mercer                      The University of Western Ontario, Canada  
Diego Mollá-Aliod                Macquarie University, Australia  
Hiromi Nakaiwa                  Nagoya University, Japan  
Tin Htar New                        Magway Computer University, Myanmar  
Fumihito Nishino                 Fujitsu Laboratories, Japan  
Win Pa Pa                          University of Computer Studies, Yangon, Myanmar

Hamman Riza	Agency for the Assessment and Application of Technology (BPPT), Indonesia
Hiroaki Saito	Keio University, Japan
Kazutaka Shimada	Kyushu Institute of Technology, Japan
Akira Shimazu	Japan Advanced Institute of Science and Technology, Japan
Kiyooki Shirai	Japan Advanced Institute of Science and Technology, Japan
Khin Mar Soe	University of Computer Studies, Yangon, Myanmar
Virach Sornlertlamvanich	Thammasat University, Thailand
Thepchai Supnithi	NECTEC, Thailand
Hisami Suzuki	Microsoft, USA
Masami Suzuki	KDDI Research, Japan
Kumiko Tanaka	The University of Tokyo, Japan
Thanaruk Theeramunkong	Thammasat University, Thailand
Aye Thida	University of Computer Studies, Mandalay, Myanmar
Takenobu Tokunaga	Tokyo Institute of Technology, Japan
Mutsuko Tomokiyo	Université Grenoble Alpes, France
Yang Xiang	University of Guelph, Canada
Yuzana	Sittwe Computer University, Myanmar
Ingrid Zukerman	Monash University, Australia



Ministry of Education,  
Myanmar



University of Computer  
Studies, Yangon,  
Myanmar

# Contents

## Semantics and Semantic Analysis

Detecting Earthquake Survivors with Serious Mental Affliction. . . . .	3
<i>Tatsuya Aoki, Katsumasa Yoshikawa, Tetsuya Nasukawa, Hiroya Takamura, and Manabu Okumura</i>	
A Deep Neural Architecture for Sentence-Level Sentiment Classification in Twitter Social Networking . . . . .	15
<i>Huy Nguyen and Minh-Le Nguyen</i>	
Learning Word Embeddings for Aspect-Based Sentiment Analysis . . . . .	28
<i>Duc-Hong Pham, Anh-Cuong Le, and Thi-Kim-Chung Le</i>	
BolLy: Annotation of Sentiment Polarity in Bollywood Lyrics Dataset . . . . .	41
<i>G. Drushti Apoorva and Radhika Mamidi</i>	
Frame-Based Semantic Patterns for Relation Extraction . . . . .	51
<i>Angrosh Mandya, Danushka Bollegala, Frans Coenen, and Katie Atkinson</i>	
Semantic Refinement GRU-Based Neural Language Generation for Spoken Dialogue Systems . . . . .	63
<i>Van-Khanh Tran and Le-Minh Nguyen</i>	
Discovering Representative Space for Relational Similarity Measurement. . . . .	76
<i>Huda Hakami, Angrosh Mandya, and Danushka Bollegala</i>	
Norms of Valence and Arousal for 2,076 Chinese 4-Character Words . . . . .	88
<i>Pingping Liu, Minglei Li, Qin Lu, and Buxin Han</i>	

## Statistical Machine Translation

Integrating Specialized Bilingual Lexicons of Multiword Expressions for Domain Adaptation in Statistical Machine Translation . . . . .	101
<i>Nasredine Semmar and Meriama Laib</i>	
Logical Parsing from Natural Language Based on a Neural Translation Model . . . . .	115
<i>Liang Li, Yifan Liu, Zengchang Qin, Pengyu Li, and Tao Wan</i>	



Phrase-Level Grouping for Lexical Gap Resolution  
in Korean-Vietnamese SMT . . . . . 127  
*Seung Woo Cho, Eui-Hyeon Lee, and Jong-Hyeok Lee*

Enhancing Pivot Translation Using Grammatical  
and Morphological Information. . . . . 137  
*Hai-Long Trieu and Le-Minh Nguyen*

**Corpora and Corpus-Based Language Processing**

Information-Structure Annotation of the “Balanced Corpus  
of Contemporary Written Japanese”. . . . . 155  
*Takuya Miyauchi, Masayuki Asahara, Natsuko Nakagawa,  
and Sachi Kato*

**Syntax and Syntactic Analysis**

Khmer POS Tagging Using Conditional Random Fields . . . . . 169  
*Sokunsatya Sangvat and Charnyote Pluempitiwiriyawej*

Statistical Khmer Name Romanization . . . . . 179  
*Chenchen Ding, Vichet Chea, Masao Utiyama, Eiichiro Sumita,  
Sethserey Sam, and Sopheap Seng*

Burmese (Myanmar) Name Romanization: A Sub-syllabic Segmentation  
Scheme for Statistical Solutions . . . . . 191  
*Chenchen Ding, Win Pa Pa, Masao Utiyama, and Eiichiro Sumita*

**Document Classification**

Domain Adaptation for Document Classification by Alternately Using  
Semi-supervised Learning and Feature Weighted Learning . . . . . 205  
*Hiroyuki Shinnou, Kanako Komiya, and Minoru Sasaki*

**Information Extraction and Text Mining**

End-to-End Recurrent Neural Network Models for Vietnamese  
Named Entity Recognition: Word-Level Vs. Character-Level . . . . . 219  
*Thai-Hoang Pham and Phuong Le-Hong*

Nested Named Entity Recognition Using Multilayer Recurrent  
Neural Networks. . . . . 233  
*Truong-Son Nguyen and Le-Minh Nguyen*

**Text Summarization**

Deletion-Based Sentence Compression Using Bi-enc-dec LSTM . . . . . 249  
*Dac-Viet Lai, Nguyen Truong Son, and Nguyen Le Minh*

**Text and Message Understanding**

Myanmar Number Normalization for Text-to-Speech . . . . . 263  
*Aye Mya Hlaing, Win Pa Pa, and Ye Kyaw Thu*

*Expect the Unexpected: Harnessing Sentence Completion*  
 for Sarcasm Detection . . . . . 275  
*Aditya Joshi, Samarth Agrawal, Pushpak Bhattacharyya,*  
*and Mark J. Carman*

Detecting Computer-Generated Text Using Fluency and Noise Features . . . . . 288  
*Hoang-Quoc Nguyen-Son and Isao Echizen*

**Automatic Speech Recognition**

Speaker Adaptation on Myanmar Spontaneous Speech Recognition . . . . . 303  
*Hay Mar Soe Naing and Win Pa Pa*

Exploring the Effect of Tones for Myanmar Language Speech Recognition  
 Using Convolutional Neural Network (CNN) . . . . . 314  
*Aye Nyein Mon, Win Pa Pa, and Ye Kyaw Thu*

**Spoken Language and Dialogue**

Listenability Measurement Based on Learners’ Transcription Performance . . . 329  
*Katsunori Kotani and Takehiko Yoshimi*

**Speech Pathology**

A Robust Algorithm for Pathological-Speech Correction . . . . . 341  
*Naim Terbeh and Mounir Zrigui*

**Speech Analysis**

Identification of Pronunciation Defects in Spoken Arabic Language . . . . . 355  
*Naim Terbeh and Mounir Zrigui*

**Author Index** . . . . . 367