7th Workshop on Asian Language Resources 2009

(ALR-7)

Held at ACL-IJCNLP 2009

Singapore
6-7 August 2009

ISBN: 978-1-61738-456-1

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© (2009) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2010)

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: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

Enhancing the Japanese WordNet
Francis Bond, Hitoshi Isahara, Sanae Fujita, Kiyotaka Uchimoto, Takayuki Kuribayashi and Kyoko Kanzaki 1

An Empirical Study of Vietnamese Noun Phrase Chunking with Discriminative Sequence Models
Le Minh Nguyen, Huong Thao Nguyen, Phuong Thai Nguyen, Tu Bao Ho and Akira Shimazu . . 9

Corpus-based Sinhala Lexicon
Ruvan Weerasinghe, Dulip Herath and Viraj Welgama 17

Analysis and Development of Urdu POS Tagged Corpus
Ahmed Muaz, Aasim Ali and Sarmad Hussain 24

Annotating Dialogue Acts to Construct Dialogue Systems for Consulting
Kiyonori Ohtake, Teruhisa Misu, Chiori Hori, Hideki Kashioka and Satoshi Nakamura 32

Assas-band, an Affix-Exception-List Based Urdu Stemmer
Qurat-ul-Ain Akram, Asma Naseer and Sarmad Hussain 40

Automated Mining Of Names Using Parallel Hindi-English Corpus
R. Mahesh K. Sinha 48

Basic Language Resources for Diverse Asian Languages: A Streamlined Approach for Resource Creation
Heather Simpson, Kazuaki Maeda and Christopher Cieri 55

Finite-State Description of Vietnamese Reduplication
Le Hong Phuong, Nguyen Thi Minh Huyen and Roussanaly Azim 63

Construction of Chinese Segmented and POS-tagged Conversational Corpora and Their Evaluations on Spontaneous Speech Recognitions
Xinhui Hu, Ryosuke Isotani and Satoshi Nakamura 70

Bengali Verb Subcategorization Frame Acquisition - A Baseline Model
Somnath Banerjee, Dipankar Das and Sivaji Bandyopadhyay 76

Phonological and Logographic Influences on Errors in Written Chinese Words
Chao-Lin Liu, Kan-Wen Tien, Min-Hua Lai, Yi-Hsuan Chuang and Shih-Hung Wu 84

Resource Report: Building Parallel Text Corpora for Multi-Domain Translation System
Budiono, Hammam Riza and Chairil Hakim 92

A Syntactic Resource for Thai: CG Treebank
Taneth Ruangrajitpakorn, Kanokorn Trakultaweekoon and Thepchai Supnithi 96

Part of Speech Tagging for Mongolian Corpus
Purev Jaimai and Odbayar Chimeddorj 103

Interaction Grammar for the Persian Language: Noun and Adjectival Phrases
Masood Ghayoomi and Bruno Guillaume 107

KTimeML: Specification of Temporal and Event Expressions in Korean Text
Seohyun Im, Hyunjo You, Hayun Jang, Seungho Nam and Hyopil Shin 115

CWN-LMF: Chinese WordNet in the Lexical Markup Framework
Lung-Hao Lee, Shu-Kai Hsieh and Chu-Ren Huang 123

Philippine Language Resources: Trends and Directions
Rachel Edita Roxas, Charibeth Cheng and Nathalie Rose Lim 131

Thai WordNet Construction
Sareewan Thoongsup, Thatsanee Charoenporn, Kergrit Robkop, Tan Sinthurahat, Chumpol Mokarat, Virach Sornlertlamvanich and Hitoshi Isahara 139

Query Expansion using LMF-Compliant Lexical Resources
Takenobu Tokunaga, Dain Kaplan, Nicoletta Calzolari, Monica Monachini, Claudia Soria, Virach Sornlertlamvanich, Thatsanee Charoenporn, Yingju Xia, Chu-Ren Huang, Shu-Kai Hsieh and Kiyoaki Shirai 145

Thai National Corpus: A Progress Report
Wirote Aroonmanakun, Kachen Tansiri and Pairit Nittayanuparp 153

The FLaReNet Thematic Network: A Global Forum for Cooperation
Nicoletta Calzolari and Claudia Soria 161

Towards Building Advanced Natural Language Applications - An Overview of the Existing Primary Resources and Applications in Nepali
Bal Krishna Bal 165

Using Search Engine to Construct a Scalable Corpus for Vietnamese Lexical Development for Word Segmentation
Doan Nguyen 171

Word Segmentation Standard in Chinese, Japanese and Korean
Key-Sun Choi, Hitoshi Isahara, Kyoko Kanzaki, Hansaem Kim, Seok Mun Pak and Maosong Sun 179