

2nd Workshop on South and Southeast Asian Natural Language Processing 2011

(WSSANLP 2011)

**Held at the 5th International Joint Conference on Natural
Language Processing 2011 (IJCNLP 2011)**

**Chiang Mai, Thailand
8 November 2011**

ISBN: 978-1-62276-441-9

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© (2011) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2012)

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

<i>Hybrid Inflectional Stemmer and Rule-based Derivational Stemmer for Gujarati</i> Kartik Suba, Dipti Jiandani and Pushpak Bhattacharyya	1
<i>Improving Persian-English Statistical Machine Translation: Experiments in Domain Adaptation</i> Mahsa Mohaghegh, Abdolhossein Sarrafzadeh and Tom Moir	9
<i>Thai Word Segmentation Verification Tool</i> Supon Klaithin, Kanyanut Kriengkiet, Sitthaa Phaholphyinyo and Krit Kosawat	16
<i>The Semi-Automatic Construction of Part-Of-Speech Taggers for Specific Languages by Statistical Methods</i> Tomohiro YAMASAKI, Hiromi WAKAKI and Masaru SUZUKI	23
<i>Towards a Malay Derivational Lexicon: Learning Affixes Using Expectation Maximization</i> Suriani Sulaiman, Michael Gasser and Sandra Kuebler	30
<i>Punjabi Language Stemmer for nouns and proper names</i> Vishal Gupta and Gurpreet Singh Lehal	35
<i>Challenges in Urdu Text Tokenization and Sentence Boundary Disambiguation</i> Zobia Rehman, Waqas Anwar and Usama Ijaz Bajwa	40
<i>Challenges in Developing a Rule based Urdu Stemmer</i> Sajjad Ahmad Khan, Waqas Anwar and Usama Ijaz Bajwa	46
<i>Developing a New System for Arabic Morphological Analysis and Generation</i> Mourad Gridach and Noureddine Chenfour	52