1st Workshop on South and Southeast Asian Natural Language Processing

At the 23rd International Conference on Computational Linguistics 2010

Beijing, China
24 August 2010

ISBN: 978-1-61782-345-9

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.

Printed by Curran Associates, Inc. (2011)

Chinese Information Processing Society of China
No. 4 Zhongguancun South 4th St.
Haidian District Beijing
China 100190

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

Conference Program

Tuesday, August 24, 2010

16:00–16:10 Opening Remarks

16:10–16:40 Invited Talk by Dr. Rajeev Sangal

16:40–17:00 *Boosting N-gram Coverage for Unsegmented Languages Using Multiple Text Segmentation Approach*
Solomon Teferra Abate, Laurent Besacier and Sopheap Seng

17:00–17:20 *Thai Sentence-Breaking for Large-Scale SMT*
Glenn Slayden, Mei-Yuh Hwang and Lee Schwartz

17:20–17:40 *Clause Identification and Classification in Bengali*
Aniruddha Ghosh, Amitava Das and Sivaji Bandyopadhyay

17:40–17:50 break

17:50–18:10 *A Paradigm-Based Finite State Morphological Analyzer for Marathi*
Mugdha Bapat, Harshada Gune and Pushpak Bhattacharyya

18:10–18:30 *Web Based Manipuri Corpus for Multiword NER and Reduplicated MWEs Identification using SVM*
Thoudam Doren Singh and Sivaji Bandyopadhyay

18:30–18:40 *A Word Segmentation System for Handling Space Omission Problem in Urdu Script*
Gurpreet Lehal

18:40–18:50 *Hybrid Stemmer for Gujarati*
Pratikkumar Patel, Kashyap Popat and Pushpak Bhattacharyya

18:50–19:00 Closing Remarks