

10th Workshop on Asian Language Resources

Held at the 24th International Conference on Computational Linguistics (COLING 2012)

**Mumbai, India
9 December 2012**

Editors:

**Rachel Edita O. Roxas
Ruvan Weerasinghe**

**Sarmad Hussain
Virach Sornlertamvanich**

ISBN: 978-1-62748-353-7

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© (2012) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2013)

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>Korean NLP2RDF Resources</i>	
YoungGyun Hahm, KyungTae Lim, Jungyeul Park, Yongun Yoon and Key-Sun Choi	1
<i>Building Large Scale Text Corpus for Tibetan Natural Language Processing by Extracting Text from Web Pages</i>	
Huidan Liu, Minghua Nuo, Jian Wu and Yeping He	11
<i>A Structured Approach for Building Assamese Corpus: Insights, Applications and Challenges</i>	
Prof. Shikhar Kr. Sarma, Himadri Bharali, Ambeswar Gogoi, Ratul Deka and Anup Kr Barman	21
<i>Corpus Building of Literary Lesser Rich Language-Bodo: Insights and Challenges</i>	
Biswajit Brahma, Anup Kr. Barman, Prof. Shikhar Kr. Sarma and Bhatima Boro	29
<i>Dependency Parsers for Persian</i>	
Mojgan Seraji, Beáta Megyesi and Joakim Nivre	35
<i>A New DOP Model for Phrase-structure Parsing of Persian Sentences</i>	
Zahra Sarabi and Morteza Analoui	45
<i>A Hybrid Dependency Parser for Bangla</i>	
Arnab Dhar, Sanjay Chatterji, Sudeshna Sarkar and Anupam Basu	55
<i>Repairing Bengali Verb Chunks for Improved Bengali to Hindi Machine Translation</i>	
Sanjay Chatterji, Nabanita Datta, Arnab Dhar, Biswanath Barik, Sudeshna Sarkar and Anupam Basu	65
<i>Domain Specific Ontology Extractor For Indian Languages</i>	
Brijesh Bhatt and Pushpak Bhattacharyya	75
<i>Constrained Hidden Markov Model for Bilingual Keyword Pairs Alignment</i>	
Denny Cahyadi, Fabien Cromieres and Sadao Kurohashi	85
<i>N-gram and Gazetteer List Based Named Entity Recognition for Urdu: A Scarce Resourced Language</i>	
Faryal Jahangir, Waqas Anwar, Usama Ijaz Bajwa and Xuan Wang	95
<i>Developing a POS tagger for Magahi: A Comparative Study</i>	
Ritesh Kumar, Bornini Lahiri and Deepak Alok	105
<i>Enhancing Lemmatization for Mongolian and its Application to Statistical Machine Translation</i>	
Chimeddorj Odbayar and Atsushi Fujii	115
<i>Translations of Ambiguous Hindi Pronouns to Possible Bengali Pronouns</i>	
Sanjay Chatterji, Sudeshna Sarkar and Anupam Basu	125