4th Workshop on Cross Lingual Information Access

At the 23rd International Conference on Computational Linguistics 2010

Beijing, China
28 August 2010

ISBN: 978-1-61782-347-3

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© (2010) by the Chinese Information Processing Society of China
All rights reserved.

Printed by Curran Associates, Inc. (2011)

For permission requests, please contact the Chinese Information Processing Society of China
at the address below.

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

Word Sense Disambiguation and IR
Pushpak Bhattacharyya 1

Multiliguality at NTCIR, and moving on ...
Tetsuya Sakai 2

Filtering news for epidemic surveillance: towards processing more languages with fewer resources
Gael Lejeune, Antoine Doucet, Roman Yangarber and Nadine Lucas 3

How to Get the Same News from Different Language News Papers
T Pattabhi R K Rao and Sobha Lalitha Devi 11

The Noisier the Better: Identifying Multilingual Word Translations Using a Single Monolingual Corpus
Reinhard Rapp, Michael Zock, Andrew Trotman and Yue Xu 16

Multi-Word Expression-Sensitive Word Alignment
Tsuyoshi Okita, Alfredo Maldonado Guerra, Yvette Graham and Andy Way 26

Co-occurrence Graph Based Iterative Bilingual Lexicon Extraction From Comparable Corpora
Diptesh Chatterjee, Sudeshna Sarkar and Arpit Mishra 35

A Voting Mechanism for Named Entity Translation in EnglishChinese Question Answering
Ling-Xiang Tang, Shlomo Geva, Andrew Trotman and Yue Xu 43

Ontology driven content extraction using interlingual annotation of texts in the OMNIA project
Achille Falaise, David Rouquet, Didier Schwab, Hervé Blanchon and Christian Boitet 52

Towards multi-lingual summarization: A comparative analysis of sentence extraction methods on English and Hebrew corpora
Marina Litvak, Mark Last, Slava Kisilevich, Daniel Keim, Hagay Lipman and Assaf Ben Gur . 61

More Languages, More MAP?: A Study of Multiple Assisting Languages in Multilingual PRF
Vishal Vachhani, Manoj Chinnakotla, Mitesh Khapra and Pushpak Bhattacharyya 70

Multilinguization and Personalization of NL-based Systems
Najeh Hajlaoui and Christian Boitet 79

Conference Program

Saturday, August 28, 2010

8:35–8:45 Opening Remarks by Sudeshna Sarkar, Min Zhang, Adam Lopez and Raghavendra Udupa

8:45–9:40 Invited Talk 1:

Word Sense Disambiguation and IR
Pushpak Bhattacharyya

Invited Talk 2

11:00–11:55 *Multiliguality at NTCIR, and moving on ...*
Tetsuya Sakai

9:40–10:05 *Filtering news for epidemic surveillance: towards processing more languages with fewer resources*
Gael Lejeune, Antoine Doucet, Roman Yangarber and Nadine Lucas

10:05–10:30 *How to Get the Same News from Different Language News Papers*
T Pattabhi R K Rao and Sobha Lalitha Devi

10:30–11:00 Morning Break

11:55–12:20 *The Noisier the Better: Identifying Multilingual Word Translations Using a Single Monolingual Corpus*
Reinhard Rapp, Michael Zock, Andrew Trotman and Yue Xu

12:20–13:50 Lunch

13:50–14:15 *Multi-Word Expression-Sensitive Word Alignment*
Tsuyoshi Okita, Alfredo Maldonado Guerra, Yvette Graham and Andy Way

14:15–14:40 *Co-occurrence Graph Based Iterative Bilingual Lexicon Extraction From Comparable Corpora*
Diptesh Chatterjee, Sudeshna Sarkar and Arpit Mishra

14:40–15:05 *A Voting Mechanism for Named Entity Translation in EnglishChinese Question Answering*
Ling-Xiang Tang, Shlomo Geva, Andrew Trotman and Yue Xu

Saturday, August 28, 2010 (continued)

15:05–15:30 *Ontology driven content extraction using interlingual annotation of texts in the OMNIA project*
Achille Falaise, David Rouquet, Didier Schwab, Hervé Blanchon and Christian Boitet

15:30–16:00 Afternoon Break

16:00–16:25 *Towards multi-lingual summarization: A comparative analysis of sentence extraction methods on English and Hebrew corpora*
Marina Litvak, Mark Last, Slava Kisilevich, Daniel Keim, Hagay Lipman and Assaf Ben Gur

16:25–16:50 *More Languages, More MAP?: A Study of Multiple Assisting Languages in Multilingual PRF*
Vishal Vachhani, Manoj Chinnakotla, Mitesh Khapra and Pushpak Bhattacharyya

16:50–17:15 *Multilinguization and Personalization of NL-based Systems*
Najeh Hajlaoui and Christian Boitet

17:15–17:25 Closing