

**Fourth SIGHAN Workshop on
Chinese Language Processing
(IJCNLP 2005)**

**Jeju Island, Korea
14-15 October 2005**

ISBN: 978-1-62748-367-4

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© (2005) 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

PREFACE	i
ORGANIZERS	ii
WORKSHOP PROGRAM	iii
<i>Detecting Segmentation Errors in Chinese Annotated Corpus</i> Chengjie Sun, Chang-Ning Huang, Xiaolong Wang and Mu Li	1
<i>Using Word-Pair Identifier to Improve Chinese Input System</i> Jia-Lin Tsai	9
<i>Chinese Deterministic Dependency Analyzer: Examining Effects of Global Features and Root Node Finder</i> Yuchang Cheng, Masayuki Asahara and Yuji Matsumoto	17
<i>Chinese Classifier Assignment Using SVMs</i> Hui Guo and Huayan Zhong	25
<i>Morphological features help POS tagging of unknown words across language varieties</i> Huihsin Tseng, Daniel Jurafsky and Christopher Manning	32
<i>Product Named Entity Recognition Based on Hierarchical Hidden Markov Model</i> Feifan Liu, Jun Zhao, Bibo Lv, Bo Xu and Hao Yu	40
<i>Chinese Sketch Engine and the Extraction of Grammatical Collocations</i> Chu-Ren Huang, Adam Kilgarriff, Yiching Wu, Chih-Ming Chiu, Simon Smith, Pavel Rychly, Ming-Hong Bai and Keh-Jiann Chen	48
<i>Word Meaning Inducing via Character Ontology: A Survey on the Semantic Prediction of Chinese Two-Character Words</i> Shu-Kai Hsieh	56
<i>Domain Specific Word Extraction from Hierarchical Web Documents: A First Step Toward Building Lexicon Trees from Web Corpora</i> Jing-Shin Chang	64
<i>Turn-Taking in Mandarin Dialogue: Interactions of Tone and Intonation</i> Gina-Anne Levow	72
<i>Learning a Log-Linear Model with Bilingual Phrase-Pair Features for Statistical Machine Translation</i> Bing Zhao and Alex Waibel	79
<i>Integrating Collocation Features in Chinese Word Sense Disambiguation</i> Wanyin Li, Qin Lu and Wenjie Li	87
<i>NIL Is Not Nothing: Recognition of Chinese Network Informal Language Expressions</i> Yunqing Xia, Kam-Fai Wong and Wei Gao	95

<i>The Robustness of Domain Lexico-Taxonomy: Expanding Domain Lexicon with CiLin</i> Chu-Ren Huang, Xiang-Bing Li and Jia-Fei Hong.....	103
<i>Some Studies on Chinese Domain Knowledge Dictionary and Its Application to Text Classification</i> Jingbo Zhu and Wenliang Chen	110
<i>Resolving Pronominal References in Chinese with the Hobbs Algorithm</i> Susan Converse	116
<i>The Second International Chinese Word Segmentation Bakeoff</i> Thomas Emerson	123
<i>Combination of Machine Learning Methods for Optimum Chinese Word Segmentation</i> Masayuki Asahara, Kenta Fukuoka, Ai Azuma, Chooi-Ling Goh, Yotaro Watanabe, Yuji Matsumoto and Takashi Tsuzuki	134
<i>Unigram Language Model for Chinese Word Segmentation</i> Aitao Chen, Yiping Zhou, Anne Zhang and Gordon Sun	138
<i>Report To BMM-based Chinese Word Segmentor with Context-based Unknown Word Identifier for the Second International Chinese word Segmentation Bakeoff</i> Jia-Lin Tsai.....	142
<i>An Example-Based Chinese Word Segmentation System for CWSB-2</i> Chunyu Kit and Xiaoyue Liu.....	146
<i>Chinese Word Segmentation in FTRD Beijing</i> Heng Li, Yuan Dong, Xinnian Mao, Haila Wang and Wu Liu.....	150
<i>Perceptron Learning for Chinese Word Segmentation</i> Yaoyong Li, Chuanjiang Miao, Kalina Bontcheva and Hamish Cunningham.....	154
<i>Data-driven Language Independent Word Segmentation Using Character-Level Information</i> Dong-Hee Lim and Seung-Shik Kang	158
<i>A Maximum Entropy Approach to Chinese Word Segmentation</i> Jin Kiat Low, Hwee Tou Ng and Wenyan Guo.....	161
<i>Description of the HKU Chinese Word Segmentation System for Sighan Bakeoff 2005</i> Guohong Fu, Kang-Kwong Luke and Percy Ping-Wai Wong.....	165
<i>A Conditional Random Field Word Segmenter for Sighan Bakeoff 2005</i> Huihsin Tseng, Pichuan Chang, Galen Andrew, Daniel Jurafsky and Christopher Manning.....	168
<i>Chinese Word Segmentation with Multiple Postprocessors in HIT-IRLab</i> Huipeng Zhang, Ting Liu, Jinshan Ma and Xiantao Liao.....	172
<i>Maximal Match Chinese Segmentation Augmented by Resources Generated from a Very Large Dictionary for Post-Processing</i> Ka-Po Chow, Andy C. Chin and Wing Fu Tsoi	176

<i>Chinese Word Segmentation based on Mixing Model</i> Wei Jiang, Jian Zhao, Yi Guan and Zhiming Xu	180
<i>Two-Phase LMR-RC Tagging for Chinese Word Segmentation</i> Tak Pang Lau and Irwin King.....	183
<i>Chinese Word Segmentation in ICT-NLP</i> ShuangLong Li	187
<i>Towards a Hybrid Model for Chinese Word Segmentation</i> Xiaofei Lu	189
<i>Chinese Word Segmentation Based On Direct Maximum Entropy Model</i> Wu-Guang Shi	193
<i>A Hybrid Approach to Chinese Segmentation around CRFs</i> Jun-sheng Zhou, Xin-yu Dai, Rui-yu Ni and Jia-jun Chen.....	196
AUTHOR INDEX.....	200