

Fifth SIGHAN Workshop on Chinese Language Processing

Co-located with COLING/ACL 2006

**Sydney, Australia
22-23 July 2006**

ISBN: 978-1-62748-366-7

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© (2006) 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	vii
Organizers	ix
Workshop Program	xi
<i>Improving Context Vector Models by Feature Clustering for Automatic Thesaurus Construction</i> Jia-Ming You and Keh-Jiann Chen	1
<i>Regional Variation of Domain-Specific Lexical Items: Toward a Pan-Chinese Lexical Resource</i> Oi Yee Kwong and Benjamin K. Tsou	9
<i>Mining Atomic Chinese Abbreviation Pairs: A Probabilistic Model for Single Character Word Recovery</i> Jing-Shin Chang and Wei-Lun Teng	17
<i>Features, Bagging, and System Combination for the Chinese POS Tagging Task</i> Fei Xia and Lap Cheung	25
<i>Semantic Analysis of Chinese Garden-Path Sentences</i> Yaohong Jin	33
<i>A Clustering Approach for Unsupervised Chinese Coreference Resolution</i> Chi-shing Wang and Grace Ngai	40
<i>Latent Features in Automatic Tense Translation between Chinese and English</i> Yang Ye, Victoria Li Fossum and Steven Abney	48
<i>Cluster-Based Language Model for Sentence Retrieval in Chinese Question Answering</i> Youzheng Wu, Jun Zhao and Bo Xu	56
<i>The Role of Lexical Resources in CJK Natural Language Processing</i> Jack Halpern	64
<i>Hybrid Models for Chinese Named Entity Recognition</i> Lishuang Li, Tingting Mao, Degen Huang and Yuansheng Yang	72
<i>Realization of the Chinese BA-construction in an English-Chinese Machine Translation System</i> Xiaohong Wu, Sylviane Cardey and Peter Greenfield	79
<i>A Hybrid Approach to Chinese Base Noun Phrase Chunking</i> Fang Xu, Chengqing Zong and Jun Zhao	87
<i>A SVM-Based Model for Chinese Functional Chunk Parsing</i> Yingze Zhao and Qiang Zhou	94
<i>Broadcast Audio and Video Bimodal Corpus Exploitation and Application</i> Yu Zou, Min Hou, Yudong Chen, Fengguo Hu and Li Fu	102
<i>The Third International Chinese Language Processing Bakeoff: Word Segmentation and Named Entity Recognition</i> Gina-Anne Levow	108

<i>Chinese Named Entity Recognition with Conditional Random Fields</i> Wenliang Chen, Yujie Zhang and Hitoshi Isahara	118
<i>France Telecom R&D Beijing Word Segmenter for Sighan Bakeoff 2006</i> Wu Liu, Heng Li, Yuan Dong, Nan He, Haitao Luo and Haila Wang	122
<i>Voting between Dictionary-Based and Subword Tagging Models for Chinese Word Segmentation</i> Dong Song and Anoop Sarkar	126
<i>BMM-Based Chinese Word Segmentor with Word Support Model for the SIGHAN Bakeoff 2006</i> Jia-Lin Tsai	130
<i>On Closed Task of Chinese Word Segmentation: An Improved CRF Model Coupled with Character Clustering and Automatically Generated Template Matching</i> Richard Tzong-Han Tsai, Hsieh-Chuan Hung, Cheng-Lung Sung, Hong-Jie Dai and Wen-Lian Hsu	134
<i>Chinese Word Segmentation with Maximum Entropy and N-gram Language Model</i> Xinhao Wang, Xiaojun Lin, Dianhai Yu, Hao Tian and Xihong Wu	138
<i>On Using Ensemble Methods for Chinese Named Entity Recognition</i> Chia-Wei Wu, Shyh-Yi Jan, Richard Tzong-Han Tsai and Wen-Lian Hsu	142
<i>Chinese Word Segmentation and Named Entity Recognition by Character Tagging</i> Kun Yu, Sadao Kurohashi, Hao Liu and Toshiaki Nakazawa	146
<i>Boosting for Chinese Named Entity Recognition</i> Xiaofeng Yu, Marine Carpuat and Dekai Wu	150
<i>Chinese Word Segmentation and Named Entity Recognition Based on a Context-Dependent Mutual Information Independence Model</i> Min Zhang, GuoDong Zhou, LingPeng Yang and DongHong Ji	154
<i>Word Segmentation and Named Entity Recognition for SIGHAN Bakeoff3</i> Suxiang Zhang, Ying Qin, Juan Wen and Xiaojie Wang	158
<i>An Improved Chinese Word Segmentation System with Conditional Random Field</i> Hai Zhao, Chang-Ning Huang and Mu Li	162
<i>Chinese Word Segmentation Using Various Dictionaries</i> Guo-Wei Bian	166
<i>Character Language Models for Chinese Word Segmentation and Named Entity Recognition</i> Bob Carpenter	169
<i>Chinese Named Entity Recognition with Conditional Probabilistic Models</i> Aitao Chen, Fuchun Peng, Roy Shan and Gordon Sun	173
<i>POC-NLW Template for Chinese Word Segmentation</i> Bo Chen, Weiran Xu, Tao Peng and Jun Guo	177
<i>Chinese Word Segmentation and Named Entity Recognition Based on Conditional Random Fields Models</i> Yuanyong Feng, Le Sun and Yuanhua Lv	181

<i>Maximum Entropy Word Segmentation of Chinese Text</i>	
Aaron J. Jacobs and Yuk Wah Wong	185
<i>A Pragmatic Chinese Word Segmentation System</i>	
Wei Jiang, Yi Guan and Xiao-Long Wang	189
<i>NetEase Automatic Chinese Word Segmentation</i>	
Xin Li and Shuaixiang Dai	193
<i>N-gram Based Two-Step Algorithm for Word Segmentation</i>	
Dong-Hee Lim, Kyu-Baek Hwang and Seung-Shik Kang	197
<i>Chinese Word Segmentation Based on an Approach of Maximum Entropy Modeling</i>	
Yan Song, Jiaqing Guo and Dongfeng Cai	201
<i>Using Part-of-Speech Reranking to Improve Chinese Word Segmentation</i>	
Mengqiu Wang and Yanxin Shi	205
<i>Description of the NCU Chinese Word Segmentation and Named Entity Recognition System for SIGHAN Bakeoff 2006</i>	
Yu-Chieh Wu, Jie-Chi Yang and Qian-Xiang Lin	209
<i>Chinese Named Entity Recognition with a Multi-Phase Model</i>	
Junsheng Zhou, Liang He, Xinyu Dai and Jiajun Chen	213
<i>Designing Special Post-Processing Rules for SVM-Based Chinese Word Segmentation</i>	
Muhua Zhu, Yilin Wang, Zhenxing Wang, Huizhen Wang and Jingbo Zhu	217
Author Index	