

**Sixth SIGHAN Workshop on  
Chinese Language Processing  
(IJCNLP 2008)**

**Hyderabad, India  
11-12 January 2008**

ISBN: 978-1-62748-365-0

## Table of Contents

Preface .....	iii
Organizers .....	v
Workshop Program .....	vii
<i>An Example-based Decoder for Spoken Language Machine Translation</i> Zhou-Jun Li, Wen-Han Chao and Yue-Xin Chen .....	1
<i>Automatic Extraction of English-Chinese Transliteration Pairs using Dynamic Window and Tokenizer</i> Chengguo Jin, Seung-Hoon Na, Dong-Il Kim and Jong-Hyeok Lee .....	9
<i>Mining Transliterations from Web Query Results: An Incremental Approach</i> Jin-Shea Kuo, Haizhou Li and Chih-Lung Lin .....	16
<i>An Effective Hybrid Machine Learning Approach for Coreference Resolution</i> Feiliang Ren and Jingbo Zhu .....	24
<i>Use of Event Types for Temporal Relation Identification in Chinese Text</i> Yuchang Cheng, Masayuki Asahara and Yuji Matsumoto .....	31
<i>Chinese Word Sense Disambiguation with PageRank and HowNet</i> Jinghua Wang, Jianyi Liu and Ping Zhang .....	39
<i>Stochastic Dependency Parsing Based on A* Admissible Search</i> Bor-shen Lin .....	45
<i>Analyzing Chinese Synthetic Words with Tree-based Information and a Survey on Chinese Morphologically Derived Words</i> Jia Lu, Masayuki Asahara and Yuji Matsumoto .....	53
<i>Which Performs Better on In-Vocabulary Word Segmentation: Based on Word or Character?</i> Zhenxing Wang, Changning Huang and Jingbo Zhu .....	61
<i>The Fourth International Chinese Language Processing Bakeoff: Chinese Word Segmentation, Named Entity Recognition and Chinese POS Tagging</i> Guangjin Jin and Xiao Chen .....	69
<i>A Two-Stage Approach to Chinese Part-of-Speech Tagging</i> Aitao Chen, Ya Zhang and Gordon Sun .....	82
<i>NOKIA Research Center Beijing Chinese Word Segmentation System for the SIGHAN Bakeoff 2007</i> Jiang Li, Rile Hu, Guohua Zhang, Yuezhong Tang, Zhanjiang Song and Xia Wang .....	86

<i>Chinese Word Segmentation and Named Entity Recognition Based on Conditional Random Fields</i> Xinnian Mao, Yuan Dong, Saike He, Sencheng Bao and Haila Wang .....	90
<i>BUPT Systems in the SIGHAN Bakeoff 2007</i> Ying Qin, Caixia Yuan, Jiashen Sun and Xiaojie Wang .....	94
<i>The Character-based CRF Segmenter of MSRA&amp;NEU for the 4th Bakeoff</i> Zhenxing Wang, Changning Huang and Jingbo Zhu .....	98
<i>Chinese NER Using CRFs and Logic for the Fourth SIGHAN Bakeoff</i> Xiaofeng Yu, Wai Lam, Shing-Kit Chan, Yiu Kei Wu and Bo Chen .....	102
<i>Unsupervised Segmentation Helps Supervised Learning of Character Tagging for Word Segmentation and Named Entity Recognition</i> Hai Zhao and Chunyu Kit .....	106
<i>An Agent-Based Approach to Chinese Word Segmentation</i> Samuel W.K. Chan and Mickey W.C. Chong .....	112
<i>Nanjing Normal University Segmenter for the Fourth SIGHAN Bakeoff</i> Xiaohe Chen, Bin Li, Junzhi Lu, Hongdong Nian and Xuri Tang .....	115
<i>Two Step Chinese Named Entity Recognition Based on Conditional Random Fields Models</i> Yuanyong Feng, Ruihong Huang and Le Sun .....	120
<i>A Morpheme-based Part-of-Speech Tagger for Chinese</i> Guohong Fu and Jonathan J. Webster .....	124
<i>Chinese Named Entity Recognition and Word Segmentation Based on Character</i> Jingzhou He and Houfeng Wang .....	128
<i>HMM and CRF Based Hybrid Model for Chinese Lexical Analysis</i> Degen Huang, Xiao Sun, Shidou Jiao, Lishuang Li, Zhuoye Ding and Ru Wan .....	133
<i>Chinese Tagging Based on Maximum Entropy Model</i> Ka Seng Leong, Fai Wong, Yiping Li and Ming Chui Dong .....	138
<i>Training a Perceptron with Global and Local Features for Chinese Word Segmentation</i> Dong Song and Anoop Sarkar .....	143
<i>A Study of Chinese Lexical Analysis Based on Discriminative Models</i> Guang-Lu Sun, Cheng-Jie Sun, Ke Sun and Xiao-Long Wang .....	147
<i>Word Boundary Token Model for the SIGHAN Bakeoff 2007</i> Jia-Lin Tsai .....	151
<i>An Improved CRF based Chinese Language Processing System for SIGHAN Bakeoff 2007</i> Xihong Wu, Xiaojun Lin, Xinhao Wang, Chunyao Wu, Yaozhong Zhang and Dianhai Yu .....	155

<i>Description of the NCU Chinese Word Segmentation and Part-of-Speech Tagging for SIGHAN Bakeoff 2007</i>	
Yu-Chieh Wu, Jie-Chi Yang and Yue-Shi Lee .....	161
<i>CRF-based Hybrid Model for Word Segmentation, NER and even POS Tagging</i>	
Zhiting Xu, Xian Qian, Yuejie Zhang and Yaqian Zhou .....	167
<i>CRFs-Based Named Entity Recognition Incorporated with Heuristic Entity List Searching</i>	
Fan Yang, Jun Zhao and Bo Zou .....	171
<i>A Chinese Word Segmentation System Based on Cascade Model</i>	
Jianfeng Zhang, Jiaheng Zheng, Hu Zhang and Hongye Tan .....	175
<i>Achilles: NiCT/ATR Chinese Morphological Analyzer for the Fourth Sighan Bakeoff</i>	
Ruiqiang Zhang and Eiichiro Sumita .....	178
Author Index .....	183