

Second CIPS-SIGHAN Joint Conference on Chinese Language Processing

(CLP 2012)

**Tianjin, China
20-21 December 2012**

ISBN: 978-1-62748-363-6

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

Keynote Speaks:

<i>QA: from Turing Test to Intelligent Information Service</i> Xiaoyan Zhu	1
<i>Linguistic foundation for NLP</i> Guodong Zhou	2

Research Papers:

<i>A Language Modeling Approach to Identifying Code-Switched Sentences and Words</i> Liang-Chih Yu, Wei-Cheng He and Wei-Nan Chien	3
<i>Semi-automatic Annotation of Chinese Word Structure</i> Jianqiang Ma, Chunyu Kit and Dale Gerdemann	9
<i>Building a Chinese Lexical Taxonomy</i> Xiaopeng Bai and Nianwen Xue	18
<i>Extending and Scaling up the Chinese Treebank Annotation</i> Xiuhong Zhang and Nianwen Xue	27

Task 1: Micro-blog word segmentation

<i>The CIPS-SIGHAN CLP 2012 Chinese Word Segmentation on MicroBlog Corpora Bakeoff</i> Huiming Duan, Zhifang Sui, Ye Tian and Wenjie Li	35
<i>Word Segmentation on Chinese Micro-Blog Data with a Linear-Time Incremental Model</i> Kaixu Zhang, Maosong Sun and Changle Zhou	41
<i>Soochow University Word Segmenter for SIGHAN 2012 Bakeoff</i> Yan Fang, Zhongqing Wang, Shoushan Li, Zhongguo Li, Richen Xu and Leixin Cai	47
<i>CRFs-Based Chinese Word Segmentation for Micro-Blog with Small-Scale Data</i> Longyue Wang, Derek F. Wong, Lidia S. Chao and Junwen Xing	51
<i>A Cascaded Approach for CIPS-SIGHAN Micro-Blog Word Segmentation Bakeoff 2012</i> Bei Shi, Xianpei Han and Le Sun	58
<i>Adapting Conventional Chinese Word Segmenter for Segmenting Micro-blog Text: Combining Rule-based and Statistic-based Approaches</i> Ning Xi, Bin Li, Guangchao Tang, Shujian Huang, Yinggong Zhao, Hao Zhou, Xinyu Dai and Jiajun Chen	63
<i>Cascaded Chinese Weibo Segmentation Based on CRFs</i> keli Zhong, xue Zhou, hangyu Li and caixia Yuan	69
<i>Rules-based Chinese Word Segmentation on MicroBlog for CIPS-SIGHAN on CLP2012</i> Jing Zhang, Degen Huang, Xia Han and Wei Wang	74
<i>Semi-supervised Chinese Word Segmentation for CLP2012</i> Saike HE, Nan HE, Songxiang CEN and Jun LU	79
<i>Micro blogs Oriented Word Segmentation System</i> Liu Yijia, Zhang Meishan, Che Wanxiang, Liu Ting and Deng Yihe	85
<i>Rules Design in Word Segmentation of Chinese Micro-Blog</i> Hao Zong, Derek F. Wong and Lidia S. Chao	90

<i>A Comparison of Chinese Word Segmentation on News and Microblog Corpora with a Lexicon Based Method</i>	
Yuxiang Jia, Hongying Zan, Ming Fan and Zhimin Wang	95
<i>A MMSM-based Hybrid Method for Chinese MicroBlog Word Segmentation</i>	
Xiao Sun, Chengcheng Li, Chenyi Tang and Jiaqi Ye	99
<i>Chinese Tweets Segmentation based on Morphemes</i>	
Chaoyue Wang and Guohong Fu	106

Task 2: Chinese personal name disambiguation

<i>The Task 2 of CIPS-SIGHAN 2012 Named Entity Recognition and Disambiguation in Chinese Bakeoff</i>	
Zhengyan He, Houfeng Wang and Sujian Li	108
<i>SIR-NERD: A Chinese Named Entity Recognition and Disambiguation System using a Two-Stage Method</i>	
Zehuan Peng, Le Sun and Xianpei Han	115
<i>A Template Based Hybrid Model for Chinese Personal Name Disambiguation</i>	
Hao Zong, Derek F. Wong and Lidia S. Chao	121
<i>Attribute based Chinese Named Entity Recognition and Disambiguation</i>	
Han Wei, Liu Guang, Mao Yuzhao and Huang Zhenni	127
<i>Chinese Name Disambiguation Based on Adaptive Clustering with the Attribute Features</i>	
Wei Tian, Xiao Pan, Zhengtao Yu, yantuan Xian and xiuzhen Yang	132
<i>Explore Chinese Encyclopedic Knowledge to Disambiguate Person Names</i>	
Jie Liu, Ruifeng Xu, Qin Lu and Jian Xu	138
<i>A Joint Chinese Named Entity Recognition and Disambiguation System</i>	
Longyue Wang, Shuo Li, Derek F. Wong and Lidia S. Chao	146
<i>Chinese Personal Name Disambiguation Based on Vector Space Model</i>	
Qing-hu FAN, Hong-ying ZAN, Yu-mei CHAI, Yu-xiang JIA and Gui-ling NIU	152

Task 3: Simplified Chinese parsing

<i>Evaluation Report of the third Chinese Parsing Evaluation: CIPS-SIGHAN-ParsEval-2012</i>	
Qiang Zhou	159
<i>Multiple TreeBanks Integration for Chinese Phrase Structure Grammar Parsing Using Bagging</i>	
Meishan Zhang, Wanxiang Che and Ting Liu	168
<i>Parsing TCT with Split Conjunction Categories</i>	
Dongchen Li and Xihong Wu	174
<i>Parsing Simplified Chinese and Traditional Chinese with Sentence Structure Grammar</i>	
Xiangli Wang	179
<i>A Simplified Chinese Parser with Factored Model</i>	
Qiuping Huang, Liangye He, Derek F. Wong and Lidia S. Chao	188
<i>Parsing TCT with a Coarse-to-fine Approach</i>	
Dongchen Li and Xihong Wu	194

Task 4: Traditional Chinese parsing

<i>Traditional Chinese Parsing Evaluation at SIGHAN Bake-offs 2012</i>	
Yuen-Hsien Tseng, Lung-Hao Lee and Liang-Chih Yu	199

<i>NEU Systems in SIGHAN Bakeoff 2012</i>	
Ji Ma, LongFei Bai, Zhuo Liu, Ao Zhang and Jingbo Zhu	206
<i>Adapting Multilingual Parsing Models to Sinica Treebank</i>	
Liangye He, Derek F. Wong and Lidia S. Chao	211
<i>Improving PCFG Chinese Parsing with Context-Dependent Probability Re-estimation</i>	
Yu-Ming Hsieh, Ming-Hong Bai, Jason S. Chang and Keh-Jiann Chen	216
<i>Sentence Parsing with Double Sequential Labeling in Traditional Chinese Parsing Task</i>	
Shih-Hung Wu, Hsien-You Hsieh and Liang-Pu Chen	222
<i>A Conditional Random Field-based Traditional Chinese Base Phrase Parser for SIGHAN Bake-off 2012 Evaluation</i>	
Yih-Ru Wang and Yuan-Fu Liao	231
<i>Hierarchical Maximum Pattern Matching with Rule Induction Approach for Sentence Parsing</i>	
Yi-Syun Tan, Yuan-Cheng Chu and Jui-Feng Yeh	237