

CIPS-SIGHAN Joint Conference on Chinese Language Processing (CLP 2010)

**Beijing, China
28-29 August 2010**

Editors:

Le Sun

Keh-Jiann Chen

ISBN: 978-1-62748-364-3

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 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

Invited Papers:

<i>Word Segmentation needs change- From a linguist's view</i> Zhendong Dong , Qiang Dong and Changling Hao	1
<i>Textual Emotion Processing From Event Analysis</i> Chu-Ren Huang, Ying Chen and Sophia Yat Mei Lee	8
<i>Classical Chinese Sentence Segmentation</i> Hen-Hsen Huang, Chuen-Tsai Sun and Hsin-Hsi Chen	15
<i>On Generalized-Topic-Based Chinese Discourse Structure</i> Rou Song, Yuru Jiang and Jingyi Wang	23
<i>Semantic Computing and Language Knowledge Bases</i> Lei Wang and Shiwen Yu	34

Research Papers:

<i>Semantic class induction and its application for a Chinese voice search system</i> Yali Li, Weiqun Xu and Yonghong Yan	46
<i>Reducing the False Alarm Rate of Chinese Character Error Detection and Correction</i> Shih-Hung Wu, Yong-Zhi Chen, Ping-che Yang, Tsun Ku and Chao-Lin Liu	54
<i>Automatic Identification of Chinese Event Descriptive Clause</i> Liou Chen and Qiang Zhou	62
<i>Bigram HMM with Context Distribution Clustering for Unsupervised Chinese Part-of-Speech tagging</i> Lidan Zhang, Kwok-Ping Chan, Chunyu Kit and Dongfeng Cai	71
<i>Mining Large-scale Parallel Corpora from Multilingual Patents: An English-Chinese example and its application to SMT</i> Bin LU, Benjamin K. Tsou, Tao Jiang, Oi Yee Kwong and Jingbo Zhu	79
<i>Studies on Automatic Recognition of Common Chinese Adverbs usages Based on Statistics Methods</i> Hongying Zan, Junhui Zhang, Xuefeng Zhu and Shiwen Yu	87
<i>Automatic Identification of Predicate Heads in Chinese Sentences</i> Xiaona Ren, Qiaoli Zhou, Chunyu Kit and Dongfeng Cai	93
<i>Selecting Optimal Feature Template Subset for CRFs</i> Xingjun Xu, Guanglu Sun, Yi Guan, Xishuang Dong and Sheng Li	99
<i>A Statistical NLP Approach for Feature and Sentiment Identification from Chinese Reviews</i> Zhen Hai, Kuiyu Chang, Qinbao Song and Jung-jae Kim	105
<i>Exploring Deep Belief Network for Chinese Relation Extraction</i> Yu Chen, Wenjie Li, Yan Liu, Dequan Zheng and Tiejun Zhao	113

<i>Exploring English Lexicon Knowledge for Chinese Sentiment Analysis</i> Yulan He, Harith Alani and Deyu Zhou	121
<i>Exploiting Social Q&A Collection in Answering Complex Questions</i> Youzheng Wu and Hisashi Kawai	129
<i>Treebank of Chinese Bible Translations</i> Andi Wu	137
<i>Using Topic Sentiment Sentences to Recognize Sentiment Polarity in Chinese Reviews</i> Jiang Yang and Min Hou	144
<i>The Method of Improving the Specific Language Focused Crawler</i> Shan-Bin Chan and Hayato Yamana	151
<i>Active Learning Based Corpus Annotation</i> Hongyan Song, Tianfang Yao, Chunyu Kit and Dongfeng Cai	159
<i>Improving Chinese Word Segmentation by Adopting Self-Organized Maps of Character N-gram</i> Chongyang Zhang, Zhigang Chen and Guoping Hu	167
<i>CMDMC: A Diachronic Digital Museum of Chinese Mandarin</i> Min Hou, Yu Zou, Yonglin Teng, Wei He, Yan Wang, Jun Liu and Jiyuan Wu	175
<i>Kazakh Segmentation System of Inflectional Affixes</i> Gulila Altenbek and xiao-long wang	183
<i>Space characters in Chinese semi-structured texts</i> Rongzhou Shen, Claire Grover and Ewan Klein	191

Bakeoff Papers:

Task 1: Chinese word segmentation

<i>The CIPS-SIGHAN CLP2010 Chinese Word Segmentation Backoff</i> Hongmei Zhao and Qiu Liu	199
<i>A Multi-layer Chinese Word Segmentation System Optimized for Out-of-domain Tasks</i> Qin Gao and Stephan Vogel	210
<i>HMM Revises Low Marginal Probability by CRF for Chinese Word Segmentation</i> Degen Huang, Deqin Tong and Yanyan Luo	216
<i>A Chinese Word Segmentation System Based on Structured Support Vector Machine Utilization of Un-labeled Text Corpus</i> Chongyang Zhang, Zhigang Chen and Guoping Hu	221
<i>Chinese Word Segmentation with Conditional Support Vector Inspired Markov Models</i> Yu-Chieh Wu, Jie-Chi Yang and	228
<i>A Boundary-Oriented Chinese Segmentation Method Using N-Gram Mutual Information</i> Ling-Xiang Tang, Shlomo Geva, Andrew Trotman and Yue Xu	234

<i>Adaptive Chinese Word Segmentation with Online Passive-Aggressive Algorithm</i>	
Wenjun Gao, Xipeng Qiu and Xuanjing Huang	240
<i>A Character-Based Joint Model for CIPS-SIGHAN Word Segmentation Bakeoff 2010</i>	
Kun Wang, Chengqing Zong and Keh-Yih Su	245
<i>Incorporating New Words Detection with Chinese Word Segmentation</i>	
Hua-Ping Zhang, Jian Gao, Qian Mo and He-Yan Huang	249
<i>High OOV-Recall Chinese Word Segmenter</i>	
Xiaoming Xu, Muhua Zhu, Xiaoxu Fei and Jingbo Zhu	252
<i>Chinese word segmentation model using bootstrapping</i>	
Baobao CHANG and Mairgup Mansur	256
<i>CRF-based Experiments for Cross-Domain Chinese Word Segmentation at CIPS-SIGHAN-2010</i>	
Xiao Qin, Liang Zong, Yuqian Wu, Xiaojun Wan and Jianwu Yang	261
<i>Term Contributed Boundary Tagging by Conditional Random Fields for SIGHAN 2010 Chinese Word Segmentation Bakeoff</i>	
Tian-Jian Jiang, Shih-Hung Liu, Cheng-Lung Sung and Wen-Lian Hsu Hsu	266
<i>Chinese Word Segmentation based on Mixing Multiple Preprocessor and CRF</i>	
jianping shen, xuan wang, hainan zhao and wenxiao zhang	270
<i>A domain adaption Word Segmenter For Sighan Backoff 2010</i>	
Jiang Guo, Wenjie Su and Yangsen Zhang	274
<i>An Double Hidden HMM and an CRF for Segmentation Tasks with Pinyin's Finals</i>	
Huixing Jiang and Zhe Dong	277
<i>Combining Character-Based and Subsequence-Based Tagging for Chinese Word Segmentation</i>	
Jiange Yu, Chuan Gu and Wenying Ge	282
Task 2: Chinese parsing	
<i>Chinese Syntactic Parsing Evaluation</i>	
Qiang Zhou and Jingbo Zhu	286
<i>Discriminative Parse Reranking for Chinese with Homogeneous and Heterogeneous Annotations</i>	
Weiwei Sun, Rui Wang and Yi Zhang	296
<i>The SAU Report for the 1st CIPS-SIGHAN-ParsEval-2010</i>	
Qiaoli Zhou, Wenjing Lang, Yingying Wang, Yan Wang and Dongfeng Cai	304
<i>Dependency Parser for Chinese Constituent Parsing</i>	
Xuezhe Ma, Xiaotian Zhang, Hai Zhao and Bao-Liang Lu	312
<i>Technical Report of the CCID System for the 2th Evaluation on Chinese Parsing</i>	
Guangfan Sun	318

<i>CRF tagging for head recognition based on Stanford parser</i> Yong Cheng, Chengjie Sun, Bingquan Liu and Lei Lin	322
<i>Trebank Conversion based Self-training Strategy for Parsing</i> Zhiguo Wang and Chengqing Zong	326
<i>A Chinese LPCFG Parser with Hybrid Character Information</i> Wenzhi Xu, Chaobo Sun and Caixia Yuan	334
<i>Complete Syntactic Analysis Bases on Multi-level Chunking</i> Zhipeng Jiang, Yu Zhao , Yi Guan, Chao Li and Sheng Li	341

Task 3: Chinese personal name disambiguation

<i>The Chinese Persons Name Diambiguation Evaluation: Exploration of Personal Name Disambiguation in Chinese News</i> Ying Chen, Peng Jin, Wenjie Li and Chu-Ren Huang	346
<i>A Multi-stage Clustering Framework for Chinese Personal Name Disambiguation</i> Huizhen Wang, Haibo Ding, Yingchao Shi, Ji Ma, Xiao Zhou and Jingbo Zhu	353
<i>Combine Person Name and Person Identity Recognition and Document Clustering for Chinese Person Name Disambiguation</i> Ruifeng Xu, Jun Xu, Xiangying Dai and Chunyu Kit	359
<i>A Pipeline Approach to Chinese Personal Name Disambiguation</i> Yang Song, Zhengyan He, Chen Chen and Houfeng Wang	366
<i>Chinese Personal Name Disambiguation: Technical Report of Natural Language Processing Lab of Xiamen University</i> Xiang Zhu, Xiaodong Shi, Ningfeng Liu, YingMei Guo and Yidong Chen	371
<i>Chinese Personal Name Disambiguation Based on Person Modeling</i> Hua-Ping Zhang, Zhi-Hua Liu, Qian Mo and He-Yan Huang	374
<i>Jumping Distance based Chinese Person Name Disambiguation</i> Yu Hong, Fei Pei, Yue-hui Yang, Jian-min Yao and Qiao-ming Zhu	379
<i>Research of People Disambiguation by Combining Multiple knowledges</i> Erlei Ma and Yuanchao Liu	383
<i>DLUT: Chinese Personal Name Disambiguation with Rich Features</i> Dongliang Wang and Degen Huang	386
<i>Person Name Disambiguation based on Topic Model</i> Jiashen Sun, Tianmin Wang, Li Li and Xing Wu	391
<i>PRIS at Chinese Language Processing</i> Zhang Jiayue, Cai Yichao, Li Si, Xu Weiran and Guo Jun	399

Task 4: Chinese word sense induction

<i>Overview of the Chinese Word Sense Induction Task at CLP2010</i>	
Le Sun , Zhenzhong Zhang and Qiang Dong	403
<i>Chinese Word Sense Induction with Basic Clustering Algorithms</i>	
Yuxiang Jia, Shiwen Yu and Zhengyan Chen	410
<i>Triplet-Based Chinese Word Sense Induction</i>	
Zhao Liu, Xipeng Qiu and Xuanjing Huang	415
<i>Word Sense Induction using Cluster Ensemble</i>	
Bichuan Zhang and Jiashen Sun	420
<i>LSTC System for Chinese Word Sense Induction</i>	
Peng Jin, Yihao Zhang and Rui Sun	428
<i>NEUNLPLab Chinese Word Sense Induction System for SIGHAN Bakeoff 2010</i>	
Hao Zhang, Tong Xiao and Jingbo Zhu	432
<i>Chinese Word Sense Induction based on Hierarchical Clustering Algorithm</i>	
KeCai, Xiaodong Shi, Yidong Chen,ZhehuangHuang andYanGao	436
<i>ISCAS: A System for Chinese Word Sense Induction Based on K-means Algorithm</i>	
Zhenzhong Zhang, Le Sun and Wenbo Li	441
<i>Soochow University: Description and Analysis of the Chinese Word Sense Induction System for CLP2010</i>	
Hua Xu, Bing Liu, Longhua Qian and Guodong Zhou	446
<i>K-means and Graph-based Approaches for Chinese Word Sense Induction Task</i>	
Lisha Wang, Yanzhao Dou, Xiaoling Sun and Hongfei Lin	452
<i>Applying Spectral Clustering for Chinese Word Sense Induction</i>	
Zhengyan He, Yang Song and Houfeng Wang	458