CLP 2010

CIPS-SIGHAN Joint Conference on Chinese Language Processing

Le Sun and Keh-Jiann Chen

28 – 29 August 2010
Beijing International Convention Center
Beijing, China

Production and Manufacturing by
Chinese Information Processing Society of China
All rights reserved for hard copy production.
No.4 Zhongguancun South 4th Street
Haidian District, Beijing, China

To order hard copies of this proceedings, please contact:

Mail Order Division, Chinese Information Processing Society of China
No.4 Zhongguancun South 4th Street
Haidian District, Beijing, China
Tel: +86-010-62562961
cips@iscas.ac.cn

Preface

With the rapid of expansion of Chinese language materials on the Internet, the use of natural language technology as a way of harnessing Chinese language content is drawing growing interest from researchers around the globe. The rise of China as a global power with increasing influence on the world stage is only fanning this interest. The Chinese language also has a number of characteristics that make Chinese language processing particularly challenging and intellectually rewarding. To meet the challenge, the first CIPS-SIGHAN Joint Conference on Chinese Language Processing (CLP2010) is organized under the auspices of CIPS (Chinese Information Processing Society of China) and SIGHAN, a Special Interest Group of the ACL.

The goal of CLP2010 is to bring together both established and aspiring researchers around the globe and provide a unified forum for them to showcase their research achievements, share their ideas, and frame research problems that are crucial in advancing the state-of-the-art in Chinese language processing.

There have been four successful international Chinese word segmentation bakeoffs sponsored by SIGHAN that have greatly advanced the state-of-the-art in this area. This year, in addition to the Chinese word segmentation task, the conference will include tasks in Chinese parsing, Chinese personal name disambiguation and Chinese word sense induction, hence attracting wider participation.

The proceedings includes 5 invited papers from senior researchers and 20 regular papers carefully reviewed and selected out of 31 submissions from different areas of Chinese language processing. The four bakeoff tasks have attracted more than 68 groups to submit their results. The proceedings also includes 4 overview papers that introduce the bakeoff tasks as well as the 44 bakeoff papers.

Last but not least, we would like to thank professors Chu-Ren Huang, Dan Jurafsky, Youqi Cao, and Chenqing Zong for initiating and proposing to hold this conference. We are also deeply indebted to the reviewers for their tireless and generous work.

We wish you all an enjoyable and thought-provoking conference.

Le Sun and Keh-Jiann Chen — CLP2010 General Co-Chairs
Qun Liu and Nianwen Xue — CLP2010 Program Co-Chairs

General chairs:

Le Sun, Institute of Software, Chinese Academy of Sciences
Keh-Jiann Chen, Institute of Information Science, Academia Sinica

Program chairs:

Qun Liu, Institute of Computing Technology, Chinese Academy of Sciences
Nianwen Xue, Brandeis University

Local arrangements chair:

Erhong Yang, Beijing Language and Culture University

Bakeoff chairs:

* Chinese Word Segmentation:

Qun Liu, Institute of Computing Technology, Chinese Academy of Sciences
Hongmei Zhao, Institute of Computing Technology, Chinese Academy of Sciences

* Chinese Parsing:

Qiang Zhou, Tsinghua University
Jingbo Zhu, North East University

* Chinese Personal Name disambiguation:

Maggie Li, The Hong Kong Polytechnic University
Chu-Ren Huang, Institute of Linguistics, Academia Sinica

* Chinese Word Sense Induction :

Le Sun, Institute of Software, Chinese Academy of Sciences
Zhendong Dong, Chinese Information Processing Society of China

Publications chair:

Tiejun Zhao, Harbin Institute of Technology

Publicity chair:

Bin Wang, Institute of Computing Technology, Chinese Academy of Sciences

Reviewers:

Pi-Chuan Chang	Wanxiang Che	Keh-Jiann Chen
Jinying Chen	Jiajun Chen	Boxing Chen
Xuanjing Huang	Heng Ji	Yumei Li
Maggie Li	Sujian Li	Hongfei Lin
Ting Liu	Qun Liu	Yang Liu
Zhanyi Liu	Yajuan Lv	Shaoping Ma
Haitao Mi	Jianyun Nie	Keh-Yih Su
Le Sun	Maosong Sun	Bing Sun
Huihsin Tseng	Xiaojun Wan	Houfeng Wang
Haifeng Wang	Xiaojie Wang	Bin Wang
Kam-Fai Wong	Yunfang Wu	Hua Wu
Fei Xia	Yunqing Xia	Deyi Xiong
Jinan Xu	Nianwen Xue	Muyun Yang
Erhong Yang	Guan Yi	Kun Yu
Dongdong Zhang	Min Zhang	Min Zhang
Weidong Zhan	Zhenzhong Zhang	Honemei Zhao
Guodong Zhou	Ming Zhou	Qiang Zhou
Jingbo Zhu	Chengqing Zong	

Location: 311B+C, 2nd floor, BICC

CLP-2010 Program

Day-1 (August 28 Saturday)

Morning				
Time		*Outline*	*Chair*	*Speaker & Title*
8:30-8:40		**Opening**	**Le Sun**	
8:40-9:00		**Invited Paper**	**Keh-Jiann Chen**	**Zhendong Dong, Qiang Dong and Changling Hao, Word Segmentation needs change**
9:00 - 10:20	9:00 - 9:20	**Overview of All tasks**	**Qun Liu**	Hongmei Zhao and Qun Liu, The CIPS-SIGHAN CLP 2010 Chinese Word Segmentation Bakeoff
	9:20 - 9:40			Qiang Zhou and Jingbo Zhu, Chinese Syntactic Parsing Evaluation
	9:40 - 10:00			Ying Chen, Peng Jin, Wenjie Li and Chu-Ren Huang, The Chinese Persons Name Disambiguation Evaluation: Exploration of Personal Name Disambiguation in Chinese News
	10:00 - 10:20			Le Sun Zhenzhong Zhang and Qiang Dong, Overview of the Chinese Word Sense Induction Task at CLP2010
10:20-10:50		**Coffee Break**		
10:50-11:10		**Invited Paper**	**Nianwen Xue**	**Chu-Ren Huang, Ying Chen, Sophia Yat Mei Lee, Textual Emotion Processing From Event Analysis**
11:10 - 12:10	11:10 - 11:25	**Bakeoff Paper: Task1**	**Hongmei Zhao**	Qin Gao and Stephan Vogel, A Multi-layer Chinese Word Segmentation System Optimized for Out-of-domain Tasks
	11:25 - 11:40			Degen Huang, Deqin Tong and Yanyan Luo, HMM Revises Low Marginal Probability by CRF for Chinese Word Segmentation
	11:40 - 11:55			Chongyang Zhang, Zhigang Chen and Guoping Hu , A Chinese Word Segmentation System Based on Structured Support Vector Machine Utilization of Unlabeled Text Corpus
	11:55 - 12:10			Yu-Chieh Wu, Jie-Chi Yang and Yue-Shi Lee, Chinese Word Segmentation with Conditional Support Vector In-spired Markov Models

12:10-12:30	**POSTER 1**		1. Yali Li, Weiqun Xu and Yonghong Yan, Semantic class induction and its application for a Chinese voice search system
			2. Shih-Hung Wu, Yong-Zhi Chen, Ping-che Yang, Tsun Ku and Chao-Lin Liu, Reducing the False Alarm Rate of Chinese Character Error Detection and Correction
			3. Ling-Xiang Tang, Shlomo Geva, Andrew Trotman and Yue Xu, A Boundary-Oriented Chinese Segmentation Method Using N-Gram Mutual Information
			4. Wenjun Gao, Xipeng Qiu and Xuanjing Huang, Adaptive Chinese Word Segmentation with Online Passive-Aggressive Algorithm
			5. Kun Wang, Chengqing Zong and Keh-Yih Su, A Character-Based Joint Model for CIPS-SIGHAN Word Segmentation Bakeoff 2010
			6. Hua-Ping Zhang, Jian Gao, Qian Mo and He-Yan Huang, Incorporating New Words Detection with Chinese Word Segmentation
			7. Xiaoming Xu, Muhua Zhu, Xiaoxu Fei and Jingbo Zhu, High OOV-Recall Chinese Word Segmenter
			8. Baobao Chang and Mairgup Mansur, Chinese word segmentation model using bootstrapping
			9. Xiao Qin, Liang Zong, Yuqian Wu, Xiaojun Wan and Jianwu Yang, CRF-based Experiments for Cross-Domain Chinese Word Segmentation at CIPS-SIGHAN-2010
			10. Tian-Jian Jiang, Shih-Hung Liu, Cheng-Lung Sung and Wen-Lian Hsu Hsu, Term Contributed Boundary Tagging by Conditional Random Fields for SIGHAN 2010 Chinese Word Segmentation Bakeoff
			11. Jianping Shen, Xuan Wang, Hainan Zhao and Wenxiao Zhang, Chinese Word Segmentation based on Mixing Multiple Preprocessor and CRF
			12. Guo Jiang, A domain adaption Word Segmenter
			13. Huixing Jiang and Zhe Dong, An Double Hidden HMM and an CRF for Segmentation Tasks with Pinyin's Finals
			14. Jiangde Yu, Chuan Gu and Wenying Ge, Combining Character-Based and Subsequence-Based Tagging for Chinese Word Segmentation
12:30-14:00	**Lunch**		

Afternoon				
14:00-14:20		**Invited Paper**	**Rou Song**	**Hen-Hsen Huang, Chuen-Tsai Sun and Hsin-Hsi Chen, Classical Chinese Sentence Segmentation**
14:20 - 16:00	14:20-14:40	**Research Papers**	**Jingbo Zhu**	Liou Chen and Qiang Zhou, Automatic Identification of Chinese Event Descriptive Clause
	14:40-15:00			Lidan Zhang and Kwok-Ping Chan, Bigram HMM with Context Distribution Clustering for Unsupervised Chinese Part-of-Speech tagging
	15:00-15:20			Bin LU, Benjamin K. Tsou, Tao Jiang, Oi Yee Kwong and Jingbo Zhu, Mining Large-scale Parallel Corpora from Multilingual Patents: An English-Chinese example and its application to SMT
	15:20-15:40			Hongying Zan, Junhui Zhang, Xuefeng Zhu and Shiwen Yu, Studies on Automatic Recognition of Common Chinese Adverb's usages Based on Statistics Methods
	15:40-16:00			Xiaona Ren, Qiaoli Zhou, Chunyu Kit and Dongfeng Cai, Automatic Identification of Predicate Heads in Chinese Sentences
16:00-16:30		**Coffee Break**		
16:30-16:50		**Invited Paper**	**Wenjie Li**	**Rou Song, Yuru Jiang and Jingyi Wang, On Generalized-Topic-Based Chinese Discourse Structure**
16:50 - 17:35	16:50-17:05	**Bakeoff Paper: Task2**	**Qiang Zhou**	Weiwei Sun, Rui Wang and Yi Zhang, Discriminative Parse Reranking for Chinese with Homogeneous and Heterogeneous Annotations
	17:05-17:20			Qiaoli Zhou, Wenjing Lang, Yingying Wang, Yan Wang and Dongfeng Cai, The SAU Report for the 1st CIPS-SIGHAN-ParsEval-2010
	17:20-17:35			Xuezhe Ma, Xiaotian Zhang, Hai Zhao and Bao-Liang Lu, Dependency Parser for Chinese Constituent Parsing
17:35 - 18:20	17:35-17:50	**Bakeoff Paper: Task3**	**Ying Chen**	Huizhen Wang, Haibo Ding, Yingchao Shi, JI Ma, Xiao Zhou and Jingbo Zhu, A Multi-stage Clustering Framework for Chinese Personal Name Disambiguation
	17:50-18:05			Ruifeng Xu, Jun Xu, Xiangying Dai and Chunyu Kit, Combine Person Name and Person Identity Recognition and Document Clustering for Chinese Person Name Disambiguation
	18:05-18:20			Yang Song, Zhengyan He, Chen Chen and Houfeng Wang, A Pipeline Approach to Chinese Personal Name Disambiguation

18:20-18:40	**POSTER 2**		1. Xingjun Xu, Guanglu Sun, Yi Guan, Xishuang Dong and Sheng Li, Selecting Optimal Feature Template Subset for CRFs
			2. Zhen Hai, Kuiyu Chang, Qinbao Song and Jung-jae Kim, A Statistical NLP Approach for Feature and Sentiment Identification from Chinese Reviews
			3. Guangfan Sun, Technical Report of the CCID System for the 2th Evaluation on Chinese Parsing
			4. Yong Cheng and Chengjie Sun, CRF tagging for head recognition based on Stanford parser
			5. Zhiguo Wang and Chengqing Zong, Treebank Conversion based Self-training Strategy for Parsing
			6. Wenzhi Xu, Chaobo Sun and Caixia Yuan, A Chinese LPCFG Parser with Hybrid Character Information
			7. ZhiPeng Jiang, Yu Zhao, Yi Guan, Chao Li and Sheng Li, Complete Syntactic Analysis Based on Multi-level Chunking
			8. Xiang Zhu, Xiaodong Shi, Ningfeng Liu, YingMei Guo and Yidong Chen, Chinese Personal Name Disambiguation: Technical Report of Natural Language Processing Lab of Xiamen University
			9. Hua-Ping Zhang, Zhi-Hua Liu, Qian Mo and He-Yan Huang, Chinese Personal Name Disambiguation Based on Person Modeling
			10. Yu Hong, Fei Pei, Yue-hui Yang, Jian-min Yao and Qiao-ming Zhu, Jumping Distance based Chinese Person Name Disambiguation
			11. Erlei Ma and Yuanchao Liu, Research of People disambiguation by combining multiple knowledges
			12. Dongliang Wang and Degen Huang, DLUT: Chinese Personal Name Disambiguation with Rich Features
			13. Jiashen Sun, Tianmin Wang, Li Li and Xing Wu, Person Name Disambiguation based on Topic Model
			14. Zhang Jiayue, Cai Yichao, Li Si, Xu Weiran and Guo Jun, PRIS at Chinese Language Processing --Chinese Personal Name Disambiguation

CLP-2010 Program

Day-2 (August 29 Sunday)

<table>
<tr><td colspan="5">Morning</td></tr>
<tr><td rowspan="5">8:30
-
10:10</td><td>8:30-8:50</td><td rowspan="5">Research Papers</td><td rowspan="5">Nianwen Xue</td><td>Yu Chen, Wenjie Li, Yan Liu, Dequan Zheng and Tiejun Zhao, Exploring Deep Belief Network for Chinese Relation Extraction</td></tr>
<tr><td>8:50-9:10</td><td>Yulan He, Harith Alani and Deyu Zhou, Exploring English Lexicon Knowledge for Chinese Sentiment Analysis</td></tr>
<tr><td>9:10-9:30</td><td>Youzheng Wu and Hisashi Kawai, Exploiting Social Q&A Collection in Answering Complex Questions</td></tr>
<tr><td>9:30-9:50</td><td>Andi Wu, Treebank of Chinese Bible Translations</td></tr>
<tr><td>9:50-10:10</td><td>Jiang Yang and Min Hou, Using Topic Sentiment Sentences to Recognize Sentiment Polarity in Chinese Reviews</td></tr>
<tr><td colspan="2">10:10-10:40</td><td colspan="3">Coffee Break</td></tr>
<tr><td colspan="2">10:40-11:00</td><td>Invited Paper</td><td>Chu-Ren Huang</td><td>Lei Wang and Shiwen Yu,
Semantic Computing and Language Knowledge Bases</td></tr>
<tr><td rowspan="3">11:00
-
11:45</td><td>11:00-11:15</td><td rowspan="3">Bakeoff Paper: Task4</td><td rowspan="3">Le Sun</td><td>Yuxiang Jia, Shiwen Yu and Zhengyan Chen, Chinese Word Sense Induction with Basic Clustering Algorithms</td></tr>
<tr><td>11:15-11:30</td><td>Zhao Liu, Xipeng Qiu and Xuanjing Huang, Triplet-Based Chinese Word Sense Induction</td></tr>
<tr><td>11:30-11:45</td><td>Bichuan Zhang and Jiashen Sun,
Word Sense Induction using Cluster Ensemble</td></tr>
</table>

11:45-12:05	**POSTER 3**		1. Shan-Bin Chan and Hayato Yamana, The Method of Improving the Specific Language Focused Crawler
			2. Hongyan Song and Tianfang Yao, Active Learning Based Corpus Annotation
			3. Chongyang Zhang, Zhigang Chen and Guoping Hu, Improving Chinese Word Segmentation by Adopting Self-Organized Maps of Character N-gram
			4. Min Hou, Yu Zou, Yonglin Teng, Wei He, Yan Wang, Jun Liu and Jiyuan Wu, CMDMC: A Diachronic Digital Museum of Chinese Mandarin
			5. Gulila Altenbek and Xiao-long Wang, Kazakh Segmentation System of Inflectional Affixes
			6. Rongzhou Shen, Claire Grover and Ewan Klein, Space characters in Chinese semi-structured texts
			7. Peng Jin, Yihao Zhang and Rui Sun, LSTC System for Chinese Word Sense Induction
			8. Hao Zhang, Tong Xiao and Jingbo Zhu, NEUNLPLab Chinese Word Sense Induction System for SIGHAN Bakeoff 2010
			9. Ke Cai, Xiaodong Shi, Yidong Chen, Zhehuang Huang and Yan Gao, Chinese Word Sense Induction based on Hierarchical Clustering Algorithm
			10. Zhenzhong Zhang, Le Sun and Wenbo Li, ISCAS: A System for Chinese Word Sense Induction Based on K-means Algorithm
			11. Hua Xu, Bing Liu, Longhua Qian and Guodong Zhou, Soochow University: Description and Analysis of the Chinese Word Sense Induction System for CLP2010
			12. Lisha Wang, Yanzhao Dou, Xiaoling Sun and Hongfei Lin, K-means and Graph-based Approaches for Chinese Word Sense Induction Task
			13. Zhengyan He, Yang Song and Houfeng Wang, Applying Spectral Clustering for Chinese Word Sense Induction
12:05-12:15	**Closing**	**Chu-Ren Huang**	

Table of Contents

Invited Papers:

Word Segmentation needs change- From a linguist's view
Zhendong Dong , Qiang Dong and Changling Hao 1

Textual Emotion Processing From Event Analysis
Chu-Ren Huang, Ying Chen and Sophia Yat Mei Lee 8

Classical Chinese Sentence Segmentation
Hen-Hsen Huang, Chuen-Tsai Sun and Hsin-Hsi Chen 15

On Generalized-Topic-Based Chinese Discourse Structure
Rou Song, Yuru Jiang and Jingyi Wang 23

Semantic Computing and Language Knowledge Bases
Lei Wang and Shiwen Yu 34

Research Papers:

Semantic class induction and its application for a Chinese voice search system
Yali Li, Weiqun Xu and Yonghong Yan 46

Reducing the False Alarm Rate of Chinese Character Error Detection and Correction
Shih-Hung Wu, Yong-Zhi Chen, Ping-che Yang, Tsun Ku and Chao-Lin Liu 54

Automatic Identification of Chinese Event Descriptive Clause
Liou Chen and Qiang Zhou 62

Bigram HMM with Context Distribution Clustering for Unsupervised Chinese Part-of-Speech tagging
Lidan Zhang, Kwok-Ping Chan, Chunyu Kit and Dongfeng Cai 71

Mining Large-scale Parallel Corpora from Multilingual Patents: An English-Chinese example and its application to SMT
Bin LU, Benjamin K. Tsou, Tao Jiang, Oi Yee Kwong and Jingbo Zhu 79

Studies on Automatic Recognition of Common Chinese Adverbs usages Based on Statistics Methods
Hongying Zan, Junhui Zhang, Xuefeng Zhu and Shiwen Yu 87

Automatic Identification of Predicate Heads in Chinese Sentences
Xiaona Ren, Qiaoli Zhou, Chunyu Kit and Dongfeng Cai 93

Selecting Optimal Feature Template Subset for CRFs
Xingjun Xu, Guanglu Sun, Yi Guan, Xishuang Dong and Sheng Li 99

A Statistical NLP Approach for Feature and Sentiment Identification from Chinese Reviews
Zhen Hai, Kuiyu Chang, Qinbao Song and Jung-jae Kim 105

Exploring Deep Belief Network for Chinese Relation Extraction
Yu Chen, Wenjie Li, Yan Liu, Dequan Zheng and Tiejun Zhao 113

Exploring English Lexicon Knowledge for Chinese Sentiment Analysis
Yulan He, Harith Alani and Deyu Zhou 121

Exploiting Social Q&A Collection in Answering Complex Questions
Youzheng Wu and Hisashi Kawai 129

Treebank of Chinese Bible Translations
Andi Wu 137

Using Topic Sentiment Sentences to Recognize Sentiment Polarity in Chinese Reviews
Jiang Yang and Min Hou 144

The Method of Improving the Specific Language Focused Crawler
Shan-Bin Chan and Hayato Yamana 151

Active Learning Based Corpus Annotation
Hongyan Song, Tianfang Yao, Chunyu Kit and Dongfeng Cai 159

Improving Chinese Word Segmentation by Adopting Self-Organized Maps of Character N-gram
Chongyang Zhang, Zhigang Chen and Guoping Hu 167

CMDMC: A Diachronic Digital Museum of Chinese Mandarin
Min Hou, Yu Zou, Yonglin Teng, Wei He, Yan Wang, Jun Liu and Jiyuan Wu 175

Kazakh Segmentation System of Inflectional Affixes
Gulila Altenbek and xiao-long wang 183

Space characters in Chinese semi-structured texts
Rongzhou Shen, Claire Grover and Ewan Klein 191

Bakeoff Papers:

Task 1: Chinese word segmentation

The CIPS-SIGHAN CLP2010 Chinese Word Segmentation Backoff
Hongmei Zhao and Qiu Liu 199

A Multi-layer Chinese Word Segmentation System Optimized for Out-of-domain Tasks
Qin Gao and Stephan Vogel 210

HMM Revises Low Marginal Probability by CRF for Chinese Word Segmentation
Degen Huang, Deqin Tong and Yanyan Luo 216

A Chinese Word Segmentation System Based on Structured Support Vector Machine Utilization of Unlabeled Text Corpus
Chongyang Zhang, Zhigang Chen and Guoping Hu 221

Chinese Word Segmentation with Conditional Support Vector Inspired Markov Models
Yu-Chieh Wu, Jie-Chi Yang and 228

A Boundary-Oriented Chinese Segmentation Method Using N-Gram Mutual Information
Ling-Xiang Tang, Shlomo Geva, Andrew Trotman and Yue Xu 234

Adaptive Chinese Word Segmentation with Online Passive-Aggressive Algorithm
Wenjun Gao, Xipeng Qiu and Xuanjing Huang 240

A Character-Based Joint Model for CIPS-SIGHAN Word Segmentation Bakeoff 2010
Kun Wang, Chengqing Zong and Keh-Yih Su 245

Incorporating New Words Detection with Chinese Word Segmentation
Hua-Ping Zhang, Jian Gao, Qian Mo and He-Yan Huang 249

High OOV-Recall Chinese Word Segmenter
Xiaoming Xu, Muhua Zhu, Xiaoxu Fei and Jingbo Zhu 252

Chinese word segmentation model using bootstrapping
Baobao CHANG and Mairgup Mansur 256

CRF-based Experiments for Cross-Domain Chinese Word Segmentation at CIPS-SIGHAN-2010
Xiao Qin, Liang Zong, Yuqian Wu, Xiaojun Wan and Jianwu Yang 261

Term Contributed Boundary Tagging by Conditional Random Fields for SIGHAN 2010 Chinese Word Segmentation Bakeoff
Tian-Jian Jiang, Shih-Hung Liu, Cheng-Lung Sung and Wen-Lian Hsu Hsu 266

Chinese Word Segmentation based on Mixing Multiple Preprocessor and CRF
jianping shen, xuan wang, hainan zhao and wenxiao zhang 270

A domain adaption Word Segmenter For Sighan Backoff 2010
Jiang Guo, Wenjie Su and Yangsen Zhang 274

An Double Hidden HMM and an CRF for Segmentation Tasks with Pinyin's Finals
Huixing Jiang and Zhe Dong 277

Combining Character-Based and Subsequence-Based Tagging for Chinese Word Segmentation
Jiangde Yu, Chuan Gu and Wenying Ge 282

Task 2: Chinese parsing

Chinese Syntactic Parsing Evaluation
Qiang Zhou and Jingbo Zhu 286

Discriminative Parse Reranking for Chinese with Homogeneous and Heterogeneous Annotations
Weiwei Sun, Rui Wang and Yi Zhang 296

The SAU Report for the 1st CIPS-SIGHAN-ParsEval-2010
Qiaoli Zhou, Wenjing Lang, Yingying Wang, Yan Wang and Dongfeng Cai 304

Dependency Parser for Chinese Constituent Parsing
Xuezhe Ma, Xiaotian Zhang, Hai Zhao and Bao-Liang Lu 312

Technical Report of the CCID System for the 2th Evaluation on Chinese Parsing
Guangfan Sun 318

CRF tagging for head recognition based on Stanford parser
Yong Cheng, Chengjie Sun, Bingquan Liu and Lei Lin 322

Treebank Conversion based Self-training Strategy for Parsing
Zhiguo Wang and Chengqing Zong 326

A Chinese LPCFG Parser with Hybrid Character Information
Wenzhi Xu, Chaobo Sun and Caixia Yuan 334

Complete Syntactic Analysis Bases on Multi-level Chunking
Zhipeng Jiang, Yu Zhao , Yi Guan, Chao Li and Sheng Li 341

Task 3: Chinese personal name disambiguation

The Chinese Persons Name Diambiguation Evaluation: Exploration of Personal Name Disambiguation in Chinese News
Ying Chen, Peng Jin, Wenjie Li and Chu-Ren Huang 346

A Multi-stage Clustering Framework for Chinese Personal Name Disambiguation
Huizhen Wang, Haibo Ding, Yingchao Shi, JI Ma, Xiao Zhou and Jingbo Zhu 353

Combine Person Name and Person Identity Recognition and Document Clustering for Chinese Person Name Disambiguation
Ruifeng Xu, Jun Xu, Xiangying Dai and Chunyu Kit 359

A Pipeline Approach to Chinese Personal Name Disambiguation
Yang Song, Zhengyan He, Chen Chen and Houfeng Wang 366

Chinese Personal Name Disambiguation: Technical Report of Natural Language Processing Lab of Xiamen University
Xiang Zhu, Xiaodong Shi, Ningfeng Liu, YingMei Guo and Yidong Chen 371

Chinese Personal Name Disambiguation Based on Person Modeling
Hua-Ping Zhang, Zhi-Hua Liu, Qian Mo and He-Yan Huang 374

Jumping Distance based Chinese Person Name Disambiguation
Yu Hong, Fei Pei, Yue-hui Yang, Jian-min Yao and Qiao-ming Zhu 379

Research of People Disambiguation by Combining Multiple knowledges
Erlei Ma and Yuanchao Liu 383

DLUT: Chinese Personal Name Disambiguation with Rich Features
Dongliang Wang and Degen Huang 386

Person Name Disambiguation based on Topic Model
Jiashen Sun, Tianmin Wang, Li Li and Xing Wu 391

PRIS at Chinese Language Processing
Zhang JIayue, Cai YIchao, Li Si, Xu Weiran and Guo Jun 399

Task 4: Chinese word sense induction

Overview of the Chinese Word Sense Induction Task at CLP2010
Le Sun , Zhenzhong Zhang and Qiang Dong. 403

Chinese Word Sense Induction with Basic Clustering Algorithms
Yuxiang Jia, Shiwen Yu and Zhengyan Chen 410

Triplet-Based Chinese Word Sense Induction
Zhao Liu, Xipeng Qiu and Xuanjing Huang 415

Word Sense Induction using Cluster Ensemble
Bichuan Zhang and Jiashen Sun. 420

LSTC System for Chinese Word Sense Induction
Peng Jin, Yihao Zhang and Rui Sun 428

NEUNLPLab Chinese Word Sense Induction System for SIGHAN Bakeoff 2010
Hao Zhang, Tong Xiao and Jingbo Zhu 432

Chinese Word Sense Induction based on Hierarchical Clustering Algorithm
KeCai, Xiaodong Shi, Yidong Chen,ZhehuangHuang andYanGao 436

ISCAS: A System for Chinese Word Sense Induction Based on K-means Algorithm
Zhenzhong Zhang, Le Sun and Wenbo Li 441

Soochow University: Description and Analysis of the Chinese Word Sense Induction System for CLP2010
Hua Xu, Bing Liu, Longhua Qian and Guodong Zhou 446

K-means and Graph-based Approaches for Chinese Word Sense Induction Task
Lisha Wang, Yanzhao Dou, Xiaoling Sun and Hongfei Lin. 452

Applying Spectral Clustering for Chinese Word Sense Induction
Zhengyan He, Yang Song and Houfeng Wang 458