

CLP 2014

**The Third CIPS-SIGHAN Joint Conference on  
Chinese Language Processing**

October 20-21, 2014  
Wuhan, 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*

## Preface

As the vast emerging web users generate big data globally, Chinese language resources develop rapidly accompanying with language-specific great interests in academic and industrial applications. CLP2014 targets to harness Chinese language content by reporting recent advancements in technical and linguistic aspects, and drawing interests of global researchers to pursue Chinese processing techniques in this big data age.

CLP2014 is the third conference jointly organized by the Chinese Language Processing Society of China (CIPS) and the ACL Special Interest Group on Chinese Language Processing (SIGHAN). The first conference, CLP-2010, was held on Aug 28-29, 2010 in Beijing, China, in conjunction with COLING 2010. The second conference, CLP2012, hosted by Tianjin University, was held on December 20-21, 2012 in Tianjin, China. CLP2014, hosted by the Central China Normal University, Wuhan, was held on Oct 20-21, 2014.

The series of CLP conference provide a platform for researchers around the world to present their research, share ideas, explore new research directions, and advance the state-of-the-art in Chinese language processing. Continuous efforts of the conference received more and more attention on its characteristics of Chinese language application and studies. The specialty of Chinese language poses challenging task for both linguist and computer scientists, and raises a conference tradition of the international bakeoffs. CLP2014 features four bakeoff tracks: Chinese word segmentation, Chinese Spelling Check, Simplified Chinese parsing, and Personal Attributes Extraction in Chinese Text.

Of the 18 submissions, CLP2014 selected 11 papers through rigorous review process. CLP2014 bakeoffs attract 18 groups to participate and submit 22 papers of both technical and overview papers.

We wish to thank everyone who showed interest and submitted a paper, all of the authors for their contributions, the members of the Program Committee for their thoughtful reviews, the invited speakers for sharing their perspectives on the topic, and all the attendees of the workshop. All of these factors contribute to a truly enriching event!

Le Sun, Chengqing Zong            *CLP2014 General Co-Chairs*  
Min Zhang, Levow, Gina-Anne   *CLP2014 Program Co-Chairs*



## Organizers

### General Chairs:

Le Sun, *Institute of Software, Chinese Academy of Sciences*  
Chengqing Zong, *Institute of Automation, Chinese Academy of Sciences*

### Program Chairs:

Min Zhang, *Soochow University*  
Levov, Gina-Anne, *University of Washington*

### Program Committee Members:

Hailong Cao, *Harbin Institute of Technology*  
Baobao Chang, *Peking University*  
Yidong Chen, *Xiamen University*  
Guohong Fu, *Heilongjiang University*  
Yanqing He, *Institute of Scientific and Technical Information of China*  
Hongying Zang, *Zhengzhou University*  
Yuexian Hou, *Tianjin University*  
Xuanjing Huang, *Fudan University*  
Shujian Huang, *Nanjing University*  
Wenbin Jiang, *Institute of Computing Technology, Chinese Academy of Sciences*  
Sujian Li, *Peking University*  
Yang Liu, *Tsinghua University*  
Zhiyuan Liu, *Tsinghua University*  
Haoliang Qi, *Heilongjiang Institute of Technology (HIT)*  
Bing Qin, *Harbin Institute of Technology*  
Likun Qiu, *Ludong University*  
Weiguang Qu, *Nanjing Normal University*  
Yanqiu Shao, *Beijing Language and Culture University*  
Xiaodong Shi, *Xiamen University*  
Wei Song, *Capital Normal University*  
Jinsong Su, *Xiamen University*  
Xiaojie Wang, *Beijing University of Posts and Telecommunications*  
Houfeng Wang, *Peking University*  
Bin Wang, *Institute of Computing Technology, Chinese Academy of Sciences*  
Mingwen Wang, *Jiangxi Normal University*  
Zhimin Wang, *Beijing City University*  
Ting Wang, *The National University of Defense Technology*  
Yunfang Wu, *Peking University*  
Yunqing Xia, *Tsinghua University*  
Tong Xiao, *Northeastern University*  
Xiaobing Zhao, *Inner Mongolia Normal University*  
Jun Xie, *Institute of Computing Technology, Chinese Academy of Sciences*  
DeyiXiong, *Soochow University*  
Jinan Xu, *Beijing Jiaotong University*  
Zhiming Xu, *Harbin Institute of Technology*  
Ruifeng Xu, *Harbin Institute of Technology*  
EndongXun, *Beijing Language and Culture University*

Erhong Yang, *Beijing Language and Culture University*  
Muyun Yang, *Harbin Institute of Technology*  
Tianfang Yao, *Shanghai Jiao Tong University*  
Hong Yu, *Soochow University*  
Weidong Zhan, *Peking University*  
Jiajun Zhang, *Institute of Automation, Chinese Academy of Sciences*  
Quan Zhang, *Institute of Acoustics, Chinese Academy of Sciences*  
Min Zhang, *Tsinghua University*  
HuaRui Zhang, *Peking University*  
Zhichang Zhang, *Northwest Normal University*  
Dakun Zhang, *Toshiba China*  
Yujie Zhang, *Beijing Jiaotong University*  
Yu Zhang, *Harbin Institute of Technology*  
Yangsen Zhang, *Beijing Information Science & Technology University*  
Jun Zhao, *Institute of Automation, Chinese Academy of Sciences*  
Tiejun Zhao, *Harbin Institute of Technology*  
Shiqi Zhao, *Baidu*  
Yanyan Zhao, *Harbin Institute of Technology*  
Qiang Zhou, *Tsinghua University*  
Yu Zhou, *Institute of Automation, Chinese Academy of Sciences*  
Jingbo Zhu, *Northeastern University*  
Yaohong Jin, *Beijing Normal University*

#### **Bakeoff Chairs:**

##### **\*Chinese Word Segmentation:**

Huiming Duan, *Peking University*  
Zhifang Sui, *Peking University*

##### **\*Chinese Spelling Check:**

Liang-Chih Yu, *Yuan Ze University*  
Lung-Hao Lee, *National Taiwan Normal University*  
Yuen-Hsien Tseng, *National Taiwan Normal University*  
Hsin-HisChen, *National Taiwan Normal University*

##### **\*Simplified Chinese Parsing:**

Qiang Zhou, *Tsinghua University*

##### **\*Personal Attributes Extraction in Chinese Text:**

Ruifeng Xu, *Harbin Institute of Technology*  
Jian Xu, *The Hong Kong Polytechnic University*

#### **Publications Chair:**

Xiangyu Duan, *Soochow University*

#### **Publicity Chair:**

Minhui Dong, *Institute for Infocomm Research, Singapore*  
Xianpei Han, *Institute of Software, Chinese Academy of Sciences*

**Local Arrangements Chair:**

Tingting He, *Central China Normal University*





## Table of Contents

### Invited Talks

<i>Research on Chinese discourse rhetorical structure representation scheme and corpus annotation</i> Guodong Zhou .....	1
<i>The Construction of language Resource and Knowledge Base for Chinese Language Computing</i> Zhifang Sui .....	2

### Research Papers

<i>COV Model and its Application in Chinese Part-of-Speech Tagging</i> Xing Fukun and Song Rou .....	3
<i>Maximum Entropy for Chinese Comma Classification with Rich Linguistic Features</i> Xiaojuan Li, Hua Yang and JiangPing Huang .....	11
<i>Improving Bilingual Lexicon Extraction Performance from Comparable Corpora via Optimizing Translation Candidate Lists</i> Shaoqi Wang, Miao Li, Zede Zhu, Zhenxin Yang and Shizhuang Weng .....	18
<i>Ch2R: A Chinese Chatter Robot for Online Shopping Guide</i> Peijie Huang, Xianmao Lin, Zeqi Lian, De Yang, Xiaoling Tang, Li Huang, Qiang Huang, Xiupeng Wu, Guisheng Wu and Xinrui Zhang .....	26
<i>Improving Chinese Sentence Polarity Classification via Opinion Paraphrasing</i> Guohong Fu, Yu He, Jiaying Song and Chaoyue Wang .....	35
<i>Problematic Situation Analysis and Automatic Recognition for Chinese Online Conversational System</i> Yang Xiang, Yaoyun Zhang, Xiaoqiang Zhou, Xiaolong Wang and Yang Qin .....	43
<i>Segment-based Fine-grained Emotion Detection for Chinese Text</i> Odbal and Zengfu Wang .....	52
<i>Bilingual Product Name Dictionary Construction Using a Two Stage Method</i> Yatian Shen and Xuanjing Huang .....	61
<i>Detection on Inconsistency of Verb Phrase in TreeBank</i> Chaoqun Duan, Dequan Zheng, Conghui Zhu, Sheng Li and Hongye Tan .....	70
<i>Local Phrase Reordering Model for Chinese-English Patent Machine Translation</i> Xiaodie Liu, Yun Zhu and Yaohong Jin .....	76
<i>Sentence Alignment of Historical Classics based on Mode Prediction and Term Translation Pairs</i> Chao Che and Xiaojun Zheng .....	85

### Bake-off Papers

<i>The CIPS-SIGHAN CLP 2014 Chinese Word Segmentation Bake-off</i> Huiming Duan, Zhifang Sui and Tao Ge .....	90
<i>Word Segmenter for Chinese Micro-blogging Text Segmentation – Report for CIPS-SIGHAN’2014 Bake-off</i> Lu Xiang, Xiaoqing Li and Yu Zhou .....	96

<i>Leveraging Rich Linguistic Features for Cross-domain Chinese Segmentation</i> Guohua Wu, Dezhu He, Keli Zhong, Xue Zhou and Caixia Yuan .....	101
<i>Personal Attributes Extraction in Chinese Text Bakeoff in CLP 2014: Overview</i> Ruifeng Xu, Shuai Wang, Feng Shi and Jian Xu .....	108
<i>Personal Attributes Extraction Based on the Combination of Trigger Words, Dictionary and Rules</i> Kailun Zhang, Mingyin Wang, Xiaoyue Cong, Fang Huang, Hongfa Xue, Lei Li and Zhiqiao Gao 114	
<i>An Introduction to BLCU Personal Attributes Extraction System</i> Dong YU, Cheng YU, Qin QU, Gongbo TANG, Chunhua LIU, Yue TIAN and Jing YI .....	120
<i>Overview of SIGHAN 2014 Bake-off for Chinese Spelling Check</i> Liang-Chih Yu, Lung-Hao Lee, Yuen-Hsien Tseng and Hsin-Hsi Chen .....	126
<i>Extended HMM and Ranking Models for Chinese Spelling Correction</i> Jinhua Xiong, Qiao Zhang, Jianpeng Hou, Qianbo Wang, Yuanzhuo Wang and Xueqi Cheng ..	133
<i>Chinese Word Spelling Correction Based on Rule Induction</i> Jui-Feng Yeh, Yun-Yun Lu, Chen-Hsien Lee, Yu-Hsiang Yu and Yong-Ting Chen .....	139
<i>Evaluation Report of the fourth Chinese Parsing Evaluation: CIPS-SIGHAN-ParsEval-2014</i> Qiang Zhou .....	146
<i>Generative CCG Parsing with OOV Prediction</i> Huijia Wu .....	153
<i>An Improved Graph Model for Chinese Spell Checking</i> Yang Xin, Hai Zhao, Yuzhu Wang and Zhongye Jia .....	157
<i>Introduction to NJUPT Chinese Spelling Check Systems in CLP-2014 Bakeoff</i> Lei Gu, Yong Wang and Xitao Liang .....	167
<i>Chinese Spelling Check System Based on Tri-gram Model</i> Qiang Huang, Peijie Huang, Xinrui Zhang, Weijian Xie, Kaiduo Hong, Bingzhou Chen and Lei Huang .....	173
<i>Introduction to BIT Chinese Spelling Correction System at CLP 2014 Bake-off</i> Min Liu, Ping Jian and Heyan Huang .....	179
<i>A mixed approach for Chinese word segmentation</i> Zhen Wang .....	186
<i>Extraction system for Personal Attributes Extraction of CLP2014</i> Zhen Wang .....	189
<i>A Study on Personal Attributes Extraction Based on the Combination of Sentences Classifications and Rules</i> Nan-chang Cheng, Cheng-qing Zong, Min Hou and Yong-lin Teng .....	192
<i>Chinese Spell Checking Based on Noisy Channel Model</i> Hsun-wen Chiu, Jian-cheng Wu and Jason S. Chang .....	202
<i>NTOU Chinese Spelling Check System in CLP Bake-off 2014</i> Wei-Cheng Chu and Chuan-Jie Lin .....	210

<i>NCTU and NTUT's Entry to CLP-2014 Chinese Spelling Check Evaluation</i> Yih-Ru Wang and Yuan-Fu Liao .....	216
<i>Chinese Spelling Error Detection and Correction Based on Language Model, Pronunciation, and Shape</i> Junjie Yu and Zhenghua Li .....	220



# Conference Program

**Saturday, October 20, 2014**

**8:30–8:40      Opening Ceremony**

**Invited Talks**

**8:40–9:35**

*Research on Chinese discourse rhetorical structure representation scheme and corpus annotation*

Guodong Zhou

**9:30–10:30**

*The Construction of language Resource and Knowledge Base for Chinese Language Computing*

Zhifang Sui

**Research Papers**

**10:45–11:10**

*COV Model and its Application in Chinese Part-of-Speech Tagging*

Xing Fukun and Song Rou

**11:10–11:35**

*Maximum Entropy for Chinese Comma Classification with Rich Linguistic Features*

Xiaojuan Li, Hua Yang and JiangPing Huang

**Saturday, October 20, 2014 (continued)**

**11:35–12:00**

*Improving Bilingual Lexicon Extraction Performance from Comparable Corpora via Optimizing Translation Candidate Lists*

Shaoqi Wang, Miao Li, Zede Zhu, Zhenxin Yang and Shizhuang Weng

**13:30–13:55**

*Ch2R: A Chinese Chatter Robot for Online Shopping Guide*

Peijie Huang, Xianmao Lin, Zeqi Lian, De Yang, Xiaoling Tang, Li Huang, Qiang Huang, Xiupeng Wu, Guisheng Wu and Xinrui Zhang

**13:55–14:20**

*Improving Chinese Sentence Polarity Classification via Opinion Paraphrasing*

Guohong Fu, Yu He, Jiaying Song and Chaoyue Wang

**14:20–14:45**

*Problematic Situation Analysis and Automatic Recognition for Chinese Online Conversational System*

Yang Xiang, Yaoyun Zhang, Xiaoqiang Zhou, Xiaolong Wang and Yang Qin

**Saturday, October 20, 2014 (continued)**

**14:45–15:10**

*Segment-based Fine-grained Emotion Detection for Chinese Text*

Odbal and Zengfu Wang

**15:35–16:00**

*Bilingual Product Name Dictionary Construction Using a Two Stage Method*

Yatian Shen and Xuanjing Huang

**16:00–16:25**

*Local Phrase Reordering Model for Chinese-English Patent Machine Translation*

Xiaodie Liu, Yun Zhu and Yaohong Jin

**Saturday, October 20, 2014 (continued)**

**16:25–16:50**

*Sentence Alignment of Historical Classics based on Mode Prediction and Term Translation Pairs*

Chao Che and Xiaojun Zheng

**Sunday, October 21, 2014**

**Bake-off Session 1 : Chinese Word Segmentation**

**8:40–9:05**

*The CIPS-SIGHAN CLP 2014 Chinese Word Segmentation Bake-off*

Huiming Duan, Zhifang Sui and Tao Ge

**9:05–9:30**

*Word Segmenter for Chinese Micro-blogging Text Segmentation – Report for CIPS-SIGHAN'2014 Bakeoff*

Lu Xiang, Xiaoqing Li and Yu Zhou

**9:30–9:55**

*Leveraging Rich Linguistic Features for Cross-domain Chinese Segmentation*

Guohua Wu, Dezhu He, Keli Zhong, Xue Zhou and Caixia Yuan



**Sunday, October 21, 2014 (continued)**

**Bake-off Session 2 : Personal Attributes Extraction in Chinese Text Task**

**9:55–10:20**

*Personal Attributes Extraction in Chinese Text Bakeoff in CLP 2014: Overview*

Ruifeng Xu, Shuai Wang, Feng Shi and Jian Xu

**10:40–11:05**

*Personal Attributes Extraction Based on the Combination of Trigger Words, Dictionary and Rules*

Kailun Zhang, Mingyin Wang, Xiaoyue Cong, Fang Huang, Hongfa Xue, Lei Li and Zhiqiao Gao

**13:30–13:55**

*An Introduction to BLCU Personal Attributes Extraction System*

Dong YU, Cheng YU, Qin QU, Gongbo TANG, Chunhua LIU, Yue TIAN and Jing YI

**Bake-off Session 3 : Chinese Spelling Check Task**

**Sunday, October 21, 2014 (continued)**

**13:55–14:20**

*Overview of SIGHAN 2014 Bake-off for Chinese Spelling Check*

Liang-Chih Yu, Lung-Hao Lee, Yuen-Hsien Tseng and Hsin-Hsi Chen

**14:20–14:45**

*Extended HMM and Ranking Models for Chinese Spelling Correction*

Jinhua Xiong, Qiao Zhang, Jianpeng Hou, Qianbo Wang, Yuanzhuo Wang and Xueqi Cheng

**15:10–15:35**

*Chinese Word Spelling Correction Based on Rule Induction*

Jui-Feng Yeh, Yun-Yun Lu, Chen-Hsien Lee, Yu-Hsiang Yu and Yong-Ting Chen

**Bake-off Session 4 : Simplified Chinese Parsing Task**

**15:35–16:00**

*Evaluation Report of the fourth Chinese Parsing Evaluation: CIPS-SIGHAN-ParsEval-2014*

Qiang Zhou

Sunday, October 21, 2014 (continued)

16:00–16:25

*Generative CCG Parsing with OOV Prediction*

Huijia Wu

**Bake-off Poster**

11:05–12:05

*An Improved Graph Model for Chinese Spell Checking*

Yang Xin, Hai Zhao, Yuzhu Wang and Zhongye Jia

*Introduction to NJUPT Chinese Spelling Check Systems in CLP-2014 Bakeoff*

Lei Gu, Yong Wang and Xitao Liang

*Chinese Spelling Check System Based on Tri-gram Model*

Qiang Huang, Peijie Huang, Xinrui Zhang, Weijian Xie, Kaiduo Hong, Bingzhou Chen and Lei Huang

*Introduction to BIT Chinese Spelling Correction System at CLP 2014 Bake-off*

Min Liu, Ping Jian and Heyan Huang

*A mixed approach for Chinese word segmentation*

Zhen Wang

*Extraction system for Personal Attributes Extraction of CLP2014*

Zhen Wang

*A Study on Personal Attributes Extraction Based on the Combination of Sentences Classifications and Rules*

Nan-chang Cheng, Cheng-qing Zong, Min Hou and Yong-lin Teng

*Chinese Spell Checking Based on Noisy Channel Model*

Hsun-wen Chiu, Jian-cheng Wu and Jason S. Chang

**Sunday, October 21, 2014 (continued)**

*NTOU Chinese Spelling Check System in CLP Bake-off 2014*

Wei-Cheng Chu and Chuan-Jie Lin

*NCTU and NTUT's Entry to CLP-2014 Chinese Spelling Check Evaluation*

Yih-Ru Wang and Yuan-Fu Liao

*Chinese Spelling Error Detection and Correction Based on Language Model, Pronunciation, and Shape*

Junjie Yu and Zhenghua Li