# 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, Chunjie 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, Chunjie 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
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

Guangfan Sun ........................................................................ 318
Task 3: Chinese personal name disambiguation

The Chinese Persons Name Disambiguation 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, Zhehuang Huang and Yan Gao ........... 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