Fourth SIGHAN Workshop on Chinese Language Processing

(IJCNLP 2005)

Jeju Island, Korea
14-15 October 2005

ISBN: 978-1-62748-367-4
# Table of Contents

**Preface** ................................................................. i

**Organizers** ............................................................ ii

**Workshop Program** .................................................... iii

*Detecting Segmentation Errors in Chinese Annotated Corpus*
  Chengjie Sun, Chang-Ning Huang, Xiaolong Wang and Mu Li ................. 1

*Using Word-Pair Identifier to Improve Chinese Input System*
  Jia-Lin Tsai ............................................................. 9

*Chinese Deterministic Dependency Analyzer: Examining Effects of Global Features and Root Node Finder*
  Yuchang Cheng, Masayuki Asahara and Yuji Matsumoto ........................ 17

*Chinese Classifier Assignment Using SVMs*
  Hui Guo and Huayyan Zhong ........................................... 25

*Morphological features help POS tagging of unknown words across language varieties*
  Huihsin Tseng, Daniel Jurafsky and Christopher Manning ...................... 32

*Product Named Entity Recognition Based on Hierarchical Hidden Markov Model*
  Feifan Liu, Jun Zhao, Bibo Lv, Bo Xu and Hao Yu .......................... 40

*Chinese Sketch Engine and the Extraction of Grammatical Collocations*
  Chu-Ren Huang, Adam Kilgarriff, Yiching Wu, Chih-Ming Chiu, Simon Smith, Pavel Rychly, Ming-Hong Bai and Keh-Jiann Chen .................. 48

*Word Meaning Inducing via Character Ontology: A Survey on the Semantic Prediction of Chinese Two-Character Words*
  Shu-Kai Hsieh ............................................................ 56

*Domain Specific Word Extraction from Hierarchical Web Documents: A First Step Toward Building Lexicon Trees from Web Corpora*
  Jing-Shin Chang .......................................................... 64

*Turn-Taking in Mandarin Dialogue: Interactions of Tone and Intonation*
  Gina-Anne Levow ........................................................ 72

*Learning a Log-Linear Model with Bilingual Phrase-Pair Features for Statistical Machine Translation*
  Bing Zhao and Alex Waibel ............................................. 79

*Integrating Collocation Features in Chinese Word Sense Disambiguation*
  Wanyin Li, Qin Lu and Wenjie Li ....................................... 87

*NIL Is Not Nothing: Recognition of Chinese Network Informal Language Expressions*
  Yunqing Xia, Kam-Fai Wong and Wei Gao .................................. 95

vi
The Robustness of Domain Lexico-Taxonomy: Expanding Domain Lexicon with CiLin
Chu-Ren Huang, Xiang-Bing Li and Jia-Fei Hong................................. 103

Some Studies on Chinese Domain Knowledge Dictionary and Its Application to Text
Classification
Jingbo Zhu and Wenliang Chen .................................................... 110

Resolving Pronominal References in Chinese with the Hobbs Algorithm
Susan Converse ........................................................................... 116

The Second International Chinese Word Segmentation Bakeoff
Thomas Emerson ................................................................. 123

Combination of Machine Learning Methods for Optimum Chinese Word Segmentation
Masayuki Asahara, Kenta Fukuda, Ai Azuma, Chooi-Ling Goh, Yotaro Watanabe,
Yuji Matsumoto and Takashi Tsuzuki ........................................... 134

Unigram Language Model for Chinese WordSegmentation
Aitao Chen, Yiping Zhou, Anne Zhang and Gordon Sun .................... 138

Report To BMM-based Chinese Word Segmentor with Context-based Unknown Word
Identifier for the Second International Chinese word Segmentation Bakeoff
Jia-Lin Tsai................................................................. 142

An Example-Based Chinese Word Segmentation System for CWSB-2
Chunyu Kit and Xiaoyue Liu ......................................................... 146

Chinese Word Segmentation in FTRD Beijing
Heng Li, Yuan Dong, Xinnian Mao, Haila Wang and Wu Liu.................. 150

Perceptron Learning for Chinese Word Segmentation
Yaoyong Li, Chuanjiang Miao, Kalina Bontcheva and Hamish Cunningham....... 154

Data-driven Language Independent Word Segmentation Using Character-Level
Information
Dong-Hee Lim and Seung-Shik Kang ............................................. 158

A Maximum Entropy Approach to Chinese Word Segmentation
Jin Kiat Low, Hwee Tou Ng and Wenyuan Guo .................................. 161

Description of the HKU Chinese Word Segmentation System for Sighan Bakeoff 2005
Guohong Fu, Kang-Kwong Luke and Percy Ping-Wai Wong...................... 165

A Conditional Random Field Word Segmenter for Sighan Bakeoff 2005
Huihsin Tseng, Pichuan Chang, Galen Andrew, Daniel Jurafsky and
Christopher Manning.................................................................... 168

Chinese Word Segmentation with Multiple Postprocessors in HIT-IRLab
Huipeng Zhang, Ting Liu, Jinshan Ma and Xiantao Liao.......................... 172

Maximal Match Chinese Segmentation Augmented by Resources Generated from a Very
Large Dictionary for Post-Processing
Ka-Po Chow, Andy C. Chin and Wing Fu Tsoi ..................................... 176
Chinese Word Segmentation based on Mixing Model
Wei Jiang, Jian Zhao, Yi Guan and Zhiming Xu ............................................. 180

Two-Phase LMR-RC Tagging for Chinese Word Segmentation
Tak Pang Lau and Irwin King ........................................................................... 183

Chinese Word Segmentation in ICT-NLP
ShuangLong Li .................................................................................................. 187

Towards a Hybrid Model for Chinese Word Segmentation
Xiaofei Lu ......................................................................................................... 189

Chinese Word Segmentation Based On Direct Maximum Entropy Model
Wu-Guang Shi .................................................................................................... 193

A Hybrid Approach to Chinese Segmentation around CRFs
Jun-sheng Zhou, Xin-yu Dai, Rui-yu Ni and Jia-jun Chen .............................. 196

Author Index .................................................................................................... 200