

Second SIGHAN Workshop on Chinese Language Processing 2003

**Sapporo, Japan
11-12 July 2003**

ISBN: 978-1-62748-369-8

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© (2003) 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

Unsupervised Training for Overlapping Ambiguity Resolution in Chinese Word Segmentation	1
<i>Mu Li, Jianfeng Gao, Changning Huang, Jianfeng Li</i>	
Class Based Sense Definition Model for Word Sense Tagging and Disambiguation	8
<i>Tracy Lin, Jason S. Chang</i>	
Utterance Segmentation Using Combined Approach Based on Bi-directional N-gram and Maximum Entropy	16
<i>Ding Liu, Chengqing Zong</i>	
Two-Character Chinese Word Extraction Based on Hybrid of Internal and Contextual Measures	24
<i>Shengfen Luo, Maosong Sun</i>	
A Bottom-up Merging Algorithm for Chinese Unknown Word Extraction	31
<i>Wei-Yun Ma, Keh-Jiann Chen</i>	
The Effect of Rhythm on Structural Disambiguation in Chinese	39
<i>Honglin Sun, Dan Jurafsky</i>	
Annotating the Propositions in the Penn Chinese Treebank	47
<i>Nianwen Xue, Martha Palmer</i>	
CHINERS: A Chinese Named Entity Recognition System for the Sports Domain	55
<i>Tianfang Yao, Wei Ding, Gregor Erbach</i>	
Chinese Lexical Analysis Using Hierarchical Hidden Markov Model	63
<i>Hua-Ping Zhang, Qun Liu, Xue-Qi Cheng, Hao Zhang, Hong-Kui Yu</i>	
Modeling of Long Distance Context Dependency in Chinese	71
<i>Guodong Zhou</i>	
A Chinese Efficient Analyser Integrating Word Segmentation, Part-Of-Speech Tagging, Partial Parsing and Full Parsing	78
<i>Guodong Zhou, Jian Su</i>	
Building A Large Chinese Corpus Annotated With Semantic Dependency	84
<i>Mingqin Li, Juanzi Li, Zhendong Dong, Zuoying Wang, Dajin Lu</i>	
News-Oriented Automatic Chinese Keyword Indexing	92
<i>Sujian Li, Houfeng Wang, Shiwen Yu, Chengsheng Xin</i>	
Semantic Maps for Word Alignment in Bilingual Parallel Corpora	98
<i>Qing Ma, Yujie Zhang, Masaki Murata, Hitoshi Isahara</i>	
Abductive Explanation-based Learning Improves Parsing Accuracy and Efficiency	104
<i>Oliver Streiter</i>	
The Semantic Knowledge-base of Contemporary Chinese and its Applications in WSD	112
<i>Hui Wang, Shiwen Yu</i>	
Learning Verb-Noun Relations to Improve Parsing	119
<i>Andi Wu</i>	
Single Character Chinese Named Entity Recognition	125
<i>Xiaodan Zhu, Mu Li, Jianfeng Gao, Chang-Ning Huang</i>	
The First International Chinese Word Segmentation Bakeoff	133
<i>Richard Sproat, Thomas Emerson</i>	
Combining Segmenter and Chunker for Chinese Word Segmentation	144
<i>Masayuki Asahara, Chooi Ling Goh, Xiaojie Wang, Yuji Matsumoto</i>	
Chinese Word Segmentation Using Minimal Linguistic Knowledge	148
<i>Aitao Chen</i>	
Chinese Word Segmentation at Peking University	152
<i>Huiming Duan, Xiaojing Bai, Baobao Chang, Shiwen Yu</i>	
A Two-stage Statistical Word Segmentation System for Chinese	156
<i>Guohong Fu, K. K. Luke</i>	
Integrating Ngram Model and Case-based Learning For Chinese Word Segmentation	160
<i>Chunyu Kit, Zhiming Xu, Jonathan J. Webster</i>	
A Unicode based Adaptive Segmentor	164
<i>Q. Lu, S. T. Chan, R. F. Xu, T. S. Chiu, B. L. Li, S. W. Yu</i>	
Introduction to CKIP Chinese Word Segmentation System for the First International Chinese Word Segmentation Bakeoff	168
<i>Wei-Yun Ma, Keh-Jiann Chen</i>	
Chinese Word Segmentation in MSR-NLP	172
<i>Andi Wu</i>	

Chinese Word Segmentation as LMR Tagging	176
<i>Nianwen Xue, Libin Shen</i>	
SYSTRAN's Chinese Word Segmentation	180
<i>Jin Yang, Jean Senellart, Remi Zajac</i>	
HHMM-Based Chinese Lexical Analyzer ICTCLAS	184
<i>Hua-Ping Zhang, Hong-Kui Yu, De-Yi Xiong, Qun Liu</i>	
Chunking-based Chinese Word Tokenization	188
<i>Guodong Zhou</i>	
Author Index	