

Third SIGHAN Workshop on Chinese Language Processing 2004

Held in cooperation with ACL-2004

**Barcelona, Spain
25 July 2004**

ISBN: 978-1-62748-368-1

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© (2004) 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

<i>Segmentation of Chinese Long Sentences Using Commas</i> Meixun Jin, Mi-Young Kim, Dongil Kim and Jong-Hyeok Lee	1
<i>A Preliminary Study on Probabilistic Models for Chinese Abbreviations</i> Jing-Shin Chang and Yu-Tso Lai	9
<i>Document Re-ranking based on Global and Local Terms</i> Lingpeng Yang, DongHong Ji and Li Tang	17
<i>Adaptive Compression-based Approach for Chinese Pinyin Input</i> JinHu Huang and David Powers	24
<i>An Enhanced Model for Chinese Word Segmentation and Part-of-Speech Tagging</i> Feng Jiang, Hui Liu, Yuquan Chen and Ruzhan Lu	28
<i>Character-Sense Association and Compounding Template Similarity: Automatic Semantic Classification of Chinese Compounds</i> Chao-Jan Chen	33
<i>Chinese Chunking with Another Type of Spec</i> Hongqiao Li, Changning Huang, Jianfeng Gao and Xiaozhong Fan	41
<i>Combining Neural Networks and Statistics for Chinese Word Sense Disambiguation</i> Zhimao Lu, Ting Liu and Sheng Li	49
<i>Chinese Word Segmentation by Classification of Characters</i> Chooi-Ling GOH, Masayuki Asahara and Yuji Matsumoto	57
<i>Automated Alignment and Extraction of Bilingual Domain Ontology for Medical Domain Web Search</i> Jui-Feng Yeh, Chung-Hsiwn Wu, Ming-Jun Chen and Liang-chih Yu	65
<i>A Statistical Model for Hangeul-Hanja Conversion in Terminology Domain</i> Jin-Xia Huang, Sun-Mee Bae and Key-sun Choi	72
<i>Chinese Term Extraction from Web Pages Based on Compound Term Productivity</i> Hiroschi Nakagawa, Hiroyuki Kojima and Akira Maeda	79
<i>Using Synonym Relations in Chinese Collocation Extraction</i> Wanyin Li, Qin Lu and Ruifeng Xu	86
<i>The Construction of A Chinese Shallow Treebank</i> Ruifeng Xu, Qin Lu, Yin Li and Wanyin Li	94
<i>Combining Prosodic and Text Features for Segmentation of Mandarin Broadcast News</i> Gina-Anne Levow	102
<i>Automatic Semantic Role Assignment for a Tree Structure</i> Jia-Ming You and Keh-Jiann Chen	109
<i>A Large-Scale Semantic Structure for Chinese Sentences</i> Li Tang, Donghong Ji and Lingpeng Yang	116
<i>Do We Need Chinese Word Segmentation for Statistical Machine Translation?</i> Jia Xu, Richard Zens and Hermann Ney	122
<i>A Semi-Supervised Approach to Build Annotated Corpus for Chinese Named Entity Recognition</i> Xiaoshan Fang, Jianfeng Gao and HuanYe Sheng	129

<i>A New Chinese Natural Language Understanding Architecture Based on Multilayer Search Mechanism</i> Wanxiang Che, Ting Liu and Sheng Li	134
<i>Aligning Bilingual Corpora Using Sentences Location Information</i> Weigang Li, Ting Liu, Zhen Wang and Sheng Li	141
<i>An Integrated Method for Chinese Unknown Word Extraction</i> Zhiyong Luo and Rou Song	148