

Lecture Notes in Artificial Intelligence 3248

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Keh-Yih Su Jun'ichi Tsujii  
Jong-Hyeok Lee Oi Yee Kwong (Eds.)

# Natural Language Processing – IJCNLP 2004

First International Joint Conference  
Hainan Island, China, March 22-24, 2004  
Revised Selected Papers

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA  
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Keh-Yih Su  
Behavior Design Corporation  
2F, No.5, Industry E. Rd., IV Science-Based Industrial Park Hsinchu, Taiwan  
E-mail: kysu@bdc.com.tw

Jun'ichi Tsujii  
University of Tokyo  
Faculty of Science and Graduate School of Information Science and Technology  
Department of Computer Science  
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan  
E-mail: tsujii@is.s.u-tokyo.ac.jp

Jong-Hyeok Lee  
Pohang University of Science and Technology (POSTECH)  
Department of Computer Science and Engineering  
San 31 Hyoja-dong, Nam-gu, Pohang 790-784, Republic of Korea  
E-mail: jhlee@postech.ac.kr

Oi Yee Kwong  
City University of Hong Kong  
Language Information Sciences Research Centre  
Tat Chee Avenue, Kowloon, Hong Kong  
E-mail: rlolivia@cityu.edu.hk

Library of Congress Control Number: 2004117792

CR Subject Classification (1998): I.2.7, I.2, F.4.3, I.7, J.5, H.3, F.2

ISSN 0302-9743  
ISBN 3-540-24475-1 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media  
springeronline.com

© Springer-Verlag Berlin Heidelberg 2005  
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Olgun Computergrafik  
Printed on acid-free paper SPIN: 11382423 06/3142 5 4 3 2 1 0

# Foreword

IJCNLP 2004 heralded a new era for computational linguistics and natural language processing in Asia, as it coincided with the launching of the Asian Federation of NLP Associations (AFNLP) in Sanya, Hainan, China. This was a timely development for the most populous and linguistically diversified region in the world, and there is little wonder that this conference managed to draw over 200 participants. Following an enthusiastic response to the call for papers from 19 regions and the vigorous and rigorous efforts of the Program Committee Co-chairs, the sizable collection of papers presented at this conference enabled AFNLP to be compared with related organizations such as COLING and ACL, which have been in existence for more than 40 years.

The organizers recognized that it is important for serious attempts to be made to increase and improve efforts in this area all across Asia. Towards this end, and to contribute to a better awareness of topical issues, IJCNLP 2004 included two panels of special interest: (1) Panel on Emerging Asian Language Processing Efforts, and (2) Panel on Multilingual NLP for Public Information Services, which was devoted to the NLP requirements envisioned by the organizers of the 2008 Olympics. There were also two thematic sessions on (a) Natural Language Technology in Mobile Information Retrieval and Text Processing User Interfaces, and (b) Text Mining in Biomedicine. The offerings of this conference were further enriched by tutorials and workshops as well as an Asian Symposium on Natural Language Processing to Overcome Language Barriers, organized as a satellite event by the Institute of Electronics, Information and Communication Engineers (IEICE) and the Communications Research Laboratory (CRL) of Japan.

The publication of the collection of papers from IJCNLP 2004 by Springer in the Lecture Notes in Artificial Intelligence (LNAI) series is a welcome move and further testifies to the standards of the research efforts reported here.

A conference of this size and scope is not possible without the contributions of many individuals. Special thanks must go to the International Advisory Committee (Nicoletta Calzolari, Eva Hajicova, Eduard Hovy, Mark Johnson, Aravind Joshi, Martin Kay, Bente Maegaard, Joseph Mariani, Makoto Nagao, Donia Scott, Hozumi Tanaka, Hans Uszkoreit, Tianshun Yao), and the Organizing Committee. In particular, we are especially grateful to the Program Committee Co-chairs, Keh-Yih Su and Jun'ichi Tsujii; Publicity Chair, Hitoshi Isahara; Satellite Events Chair, Key-Sun Choi; Finance Committee Chair and Vice Chair, Kam-Fai Wong and Tom Lai, who had to deal with a complex "one conference, two currencies" situation and separate budgeting; and the Publication Chair and Vice Chairs, Jong-Hyeok Lee, Olivia Kwong and Sadao Kurohashi. The Local Organizing Committee in Beijing undertook the challenging responsibility for local arrangements in Sanya, and we are grateful to its Chair, Youqi Cao, and its Vice Chair and External Liaison Person, Maosong Sun.

We must also thank Makoto Nagao, the Honorary Chair of the conference, and the three keynote speakers, Ching-Chun Hsieh, Mark Johnson and Hans Uszkoreit.

IJCNLP 2004 must gratefully acknowledge financial sponsorship from a number of sources: the Association for Computational Linguistics (ACL), the Association for Computational Linguistics and Chinese Language Processing (ACLCLP), the Association of Natural Language Processing (ANLP), the Chinese Information Processing Society of China (CIPS), the Korean NLP Society, Microsoft Japan for the workshop “Beyond Shallow Analyses – Formalisms and Statistical Modeling for Deep Analyses,” and the Chinese Language Computer Society (CLCS) for the workshop “Named Entity Recognition for Natural Language Processing Applications.” Through CIPS, additional contributions were obtained from the National Natural Science Foundation of China, Fujitsu R&D Center Co., Ltd., Institute of Automation, Chinese Academy of Sciences, Microsoft Research Asia, NEC Laboratories, China, Peking University, TRS Information Technology Limited, and Tsinghua University. Without such support this conference would not have been possible. We also wish to thank the Finance Office of the City University of Hong Kong for graciously taking on the responsibility as credit card clearinghouse, and the City University Press and the University Printing Office for advice. Last but not least, to the many unnamed individuals and students in Beijing, Hong Kong, Pusan, Taipei and Tokyo, whose contributions were no less important in making this conference possible, we say thank you.

The Sanya region of Hainan Island has gained considerable prominence in recent years, the latest global event being the Miss World Contest. It is hoped that in future years, reference will be made in NLP circles to AFNLP’s IJCNLP in Sanya as a watershed in NLP development.

August 2004

Benjamin K. Tsou  
President, AFNLP

# Preface

This book is a collection of selected papers presented at the 1st International Joint Conference on Natural Language Processing (IJCNLP 2004), the flagship conference organized by the Asian Federation of Natural Language Processing Associations (AFNLP) which took place on Hainan Island, China during 22–24 March, 2004.

A total of 211 papers from 19 regions were submitted to the conference. The program committee which consisted of 20 internationally well-respected experts acting as area chairs and organizers of the thematic sessions, selected 66 papers for oral presentation and 35 for poster presentation. Authors were invited to revise their papers to reflect the vigorous and constructive discussions at the conference, and we received 84 revised papers for this volume.

The articles in this volume cover diverse fields in NLP, from theoretical contributions in finite state technology, parsing, discourse modeling, machine learning, etc., to basic NLP technologies of semantic disambiguation, shallow parsing, sentence/text generation, etc., to NLP applications such as machine translation, text mining and information retrieval. We also have articles on topics crucial for Asian language processing, including word segmentation, chunkers, etc. The linguistic diversity in Asia has posed exciting challenges, not only to researchers who believe in the universality of human language and the technologies in computational linguistics, but also to those interested in practical applications of NLP for Asian languages. This volume should give them a good collection of the current and the most advanced results in such research.

At the same time, since the conference attracted contributions from active and respected researchers all over the world, the articles in this volume not only showcase language-specific research output, but also represent the general current achievements in our field, from both the theoretical and technological fronts. In particular, the two thematic sessions, “Text Mining in Biomedicine” and “Natural Language Technology in Mobile Information Retrieval and Text Processing User Interfaces,” solicited high-quality contributions on these emerging technological challenges.

We would like to thank the area chairs and the reviewers who carried out professionally rigorous reviews and selection of papers for the conference, as well as the authors who revised their already high-quality papers to reflect the discussion at the conference. Special thanks are also given to Jin-Dong Kim (University of Tokyo) and Christine Günther (Springer) who gave technical and editorial support in the preparation of this volume.

August 2004

Keh-Yih Su, Jun'ichi Tsujii,  
Jong-Hyeok Lee, Oi Yee Kwong

## Conference Co-chairs

Guangnan Ni                      Chinese Academy of Engineering, Beijing  
Benjamin K. Tsou                City University of Hong Kong, Hong Kong

## Program Co-chairs

Keh-Yih Su                        Behavior Design Corporation, Hsinchu  
Jun'ichi Tsujii                  University of Tokyo, Tokyo

## Program Committee

John Carroll                      Sussex University, Brighton  
Keh-Jiann Chen                 Academia Sinica, Taipei  
Ann Copestake                 University of Cambridge, Cambridge  
Robert Dale                     Macquarie University, Sydney  
Chu-Ren Huang                 Academia Sinica, Taipei  
Chin-Hui Lee                    Georgia Institute of Technology, Atlanta  
Gary Geunbae Lee               Pohang University of Science and Technology, Pohang  
Alessandro Lenci                University of Pisa, Pisa  
Dekang Lin                      University of Alberta, Alberta  
Yuji Matsumoto                 NAIST, Nara  
Sung Hyon Myaeng               Information and Communications University, Daejeon  
Hwee Tou Ng                    National University of Singapore, Singapore  
Franz Och                        ISI-USC, Los Angeles  
Manabu Okumura                Tokyo Institute of Technology, Tokyo  
Martha Palmer                  University of Pennsylvania, Philadelphia  
Laurent Romary                 LORIA, Paris  
Declerck Thierry                DFKI and Saarland University, Saarbruecken  
Kenji Yamanishi                NEC, Tokyo  
Tiejun Zhao                     Harbin Institute of Technology, Harbin  
Chengqing Zong                 Chinese Academy of Sciences, Beijing

## Reviewers

Steven Abney	Pierrette Bouillon	Jason Chang
Hani Abu-Salem	Thorsten Brants	Eugene Charniak
Eneko Agirre	Elizabeth Owen Bratt	Hsin-Hsi Chen
Akiko Aizawa	Paul Buitelaar	John Chen
Collin Baker	Harry Bunt	Hua Cheng
Nuria Bel	Miriam Butt	Colin Cherry
Emily Bender	Jean Caelen	Lee-Feng Chien
Christian Boitet	Nicoletta Calzolari	Tat-Seng Chua
Kalina Bontcheva	Jeongwon Cha	Fabio Ciravegna

Fabio Crestani	Geert-Jan Kruijff	Stephan Oepen
Dan Cristea	Sadao Kurohashi	Kemal Oflazer
Dick Crouch	Kui-Lam Kwok	Miles Osborne
Chris Culy	Oi Yee Kwong	Kyonghee Paik
Hamish Cunningham	Tom Lai	Patrick Pantel
Walter Daelemans	Wai Lam	Jong Park
Ido Dagan	Philippe Langlais	Seong-Bae Park
Jianwu Dang	Mirella Lapata	John Patrick
Jean-Louis Dessalles	Gianni Lazzari	Adam Pease
Laurence Devillers	Gary Geunbae Lee	Ted Pedersen
Bonnie Dorr	Jong-Hyeok Lee	Gerald Penn
Phil Edmonds	Tan Lee	Wim Peters
Gregor Erbach	Mun-Kew Leong	Detlef Prescher
Christiane Fellbaum	Gina-Anne Levow	Dragomir Radev
Sandiway Fong	David Lewis	Stefan Riezler
Anette Frank	Haizhou Li	German Rigau
Pascale Fung	Hang Li	Hae-Chang Rim
Jianfeng Gao	Chin-Yew Lin	Brian Roark
Sheng Gao	Charles X. Ling	Tetsuya Sakai
Ismael García Varea	Bing Liu	Susanne Salmon-Alt
Daniel Gildea	Ting Liu	Rajeev Sangal
Claire Grover	Robert Luk	Anoop Sarkar
Koiti Hasida	Qing Ma	Mark Seligman
Wen-Lian Hsu	Bernardo Magnini	Melanie Siegel
Taiyi Huang	Robert Malouf	Jacques Siroux
Kentaro Inui	Inderjeet Mani	Virach
Masato Ishizaki	Daniel Marcu	Sornlertlamvanich
Donghong Ji	Stella Markantonatou	Richard Sproat
Hongyan Jing	Shigeru Masuyama	Suzanne Stevenson
Kristiina Jokinen	Yuji Matsumoto	Michael Strube
Arne Jönsson	Diana McCarthy	Tomek Strzalkowski
Min-Yen Kan	Kathleen McCoy	Jian Su
Seungshik Kang	Paul Mc Kevitt	Eiichiro Sumita
Kyoko Kanzaki	Helen Meng	Maosong Sun
Lauri Karttunen	Paola Merlo	Hisami Suzuki
John Kelleher	Robert Moore	Hideki Tanaka
Chris Khoo	Tatsunori Mori	Hozumi Tanaka
Genichiro Kikui	Masaaki Nagata	Takaaki Tanaka
Jong-Bok Kim	Chieko Nakabasami	Jianhua Tao
Jung-Hee Kim	Hwee Tou Ng	Declerck Thierry
Minkoo Kim	Grace Ngai	Takenobu Tokunaga
Alistair Knott	Jian-Yun Nie	Peter Turney
Philipp Koehn	Takashi Ninomiya	Akira Ushioda
Greg Kondrak	Cheng Niu	Takehito Utsuro
Anna Korhonen	Tadashi Nomoto	Antal van den Bosch



Josef van Genabith  
Gertjan van Noord  
Renata Vieira  
Stephan Vogel  
Haifeng Wang  
Taro Watanabe  
David Weir  
Fuliang Weng

Jan Wiebe  
Andi Wu  
Dekai Wu  
Lide Wu  
Jun Xu  
Peng Xu  
Bert Xue  
Endong Xun

Kenji Yamada  
Qiang Yang  
Jiangsheng Yu  
Min Zhang  
Shuwu Zhang  
Fang Zheng  
Guodong Zhou  
Ming Zhou

# Table of Contents

## Dialogue and Discourse

Fast Reinforcement Learning of Dialogue Policies Using Stable Function Approximation . . . . .	1
<i>Matthias Denecke, Kohji Dohsaka, and Mikio Nakano</i>	
Zero Pronoun Resolution Based on Automatically Constructed Case Frames and Structural Preference of Antecedents . . . . .	12
<i>Daisuke Kawahara and Sadao Kurohashi</i>	
Improving Noun Phrase Coreference Resolution by Matching Strings . . . . .	22
<i>Xiaofeng Yang, Guodong Zhou, Jian Su, and Chew Lim Tan</i>	
Combining Labeled and Unlabeled Data for Learning Cross-Document Structural Relationships . . . . .	32
<i>Zhu Zhang and Dragomir Radev</i>	

## FSA, Parsing Algorithms

Parsing Mixed Constructions in a Type Feature Structure Grammar . . . . .	42
<i>Jong-Bok Kim and Jaehyung Yang</i>	
Iterative CKY Parsing for Probabilistic Context-Free Grammars . . . . .	52
<i>Yoshimasa Tsuruoka and Jun'ichi Tsujii</i>	

## Information Extraction and Question Answering

Causal Relation Extraction Using Cue Phrase and Lexical Pair Probabilities . . . . .	61
<i>Du-Seong Chang and Key-Sun Choi</i>	
A Re-examination of IR Techniques in QA System . . . . .	71
<i>Yi Chang, Hongbo Xu, and Shuo Bai</i>	
A Novel Pattern Learning Method for Open Domain Question Answering . . . . .	81
<i>Yongping Du, Xuanjing Huang, Xin Li, and Lide Wu</i>	
Chinese Named Entity Recognition Based on Multilevel Linguistic Features . . . . .	90
<i>Honglei Guo, Jianmin Jiang, Gang Hu, and Tong Zhang</i>	

## Information Retrieval

Information Flow Analysis with Chinese Text.....	100
<i>Paulo Cheong, Dawei Song, Peter Bruza, and Kam-Fai Wong</i>	
Phoneme-Based Transliteration of Foreign Names for OOV Problem.....	110
<i>Wei Gao, Kam-Fai Wong, and Wai Lam</i>	
Window-Based Method for Information Retrieval.....	120
<i>Qianli Jin, Jun Zhao, and Bo Xu</i>	
Improving Relevance Feedback in Language Modeling Approach: Maximum a Posteriori Probability Criterion and Three-Component Mixture Model.....	130
<i>Seung-Hoon Na, In-Su Kang, and Jong-Hyeok Lee</i>	
BBS Based Hot Topic Retrieval Using Back-Propagation Neural Network.....	139
<i>Lan You, Yongping Du, Jiayin Ge, Xuanjing Huang, and Lide Wu</i>	
How Effective Is Query Expansion for Finding Novel Information?.....	149
<i>Min Zhang, Chuan Lin, and Shaoping Ma</i>	

## Lexical Semantics, Ontology and Linguistic Resources

The Hinoki Treebank. A Treebank for Text Understanding.....	158
<i>Francis Bond, Sanae Fujita, Chikara Hashimoto, Kaname Kasahara, Shigeko Nariyama, Eric Nichols, Akira Ohtani, Takaaki Tanaka, and Shigeaki Amano</i>	
Building a Parallel Bilingual Syntactically Annotated Corpus.....	168
<i>Jan Cuřin, Martin Čmejrek, Jiří Havelka, and Vladislav Kuboň</i>	
Acquiring Bilingual Named Entity Translations from Content-Aligned Corpora.....	177
<i>Tadashi Kumano, Hideki Kashioka, Hideki Tanaka, and Takahiro Fukusima</i>	
Visual Semantics and Ontology of Eventive Verbs.....	187
<i>Minhua Ma and Paul Mc Kevitt</i>	
A Persistent Feature-Object Database for Intelligent Text Archive Systems.....	197
<i>Takashi Ninomiya, Jun'ichi Tsujii, and Yusuke Miyao</i>	

## Machine Translation and Multilinguality

Example-Based Machine Translation Without Saying Inferable Predicate . . . . .	206
<i>Eiji Aramaki, Sadao Kurohashi, Hideki Kashioka, and Hideki Tanaka</i>	
Improving Back-Transliteration by Combining Information Sources . . . . .	216
<i>Slaven Bilac and Hozumi Tanaka</i>	
Bilingual Sentence Alignment Based on Punctuation Statistics and Lexicon . . . . .	224
<i>Thomas C. Chuang, Jian-Cheng Wu, Tracy Lin, Wen-Chie Shei, and Jason S. Chang</i>	
Automatic Learning of Parallel Dependency Treelet Pairs . . . . .	233
<i>Yuan Ding and Martha Palmer</i>	
Practical Translation Pattern Acquisition from Combined Language Resources . . . . .	244
<i>Mihoko Kitamura and Yuji Matsumoto</i>	
An English-Hindi Statistical Machine Translation System . . . . .	254
<i>Raghavendra Udupa U. and Tanveer A. Faruque</i>	

## NLP Software and Application

Robust Speaker Identification System Based on Wavelet Transform and Gaussian Mixture Model . . . . .	263
<i>Wan-Chen Chen, Ching-Tang Hsieh, and Eugene Lai</i>	
Selecting Prosody Parameters for Unit Selection Based Chinese TTS . . . . .	272
<i>Minghui Dong, Kim-Teng Lua, and Jun Xu</i>	
Natural Language Database Access Using Semi-automatically Constructed Translation Knowledge . . . . .	280
<i>In-Su Kang, Jae-Hak J. Bae, and Jong-Hyeok Lee</i>	
Korean Stochastic Word-Spacing with Dynamic Expansion of Candidate Words List . . . . .	290
<i>Mi-young Kang, Sung-ja Choi, Ae-sun Yoon, and Hyuk-chul Kwon</i>	
You Don't Have to Think Twice if You Carefully Tokenize . . . . .	299
<i>Stefan Klatt and Bernd Bohnet</i>	
Automatic Genre Detection of Web Documents . . . . .	310
<i>Chul Su Lim, Kong Joo Lee, and Gil Chang Kim</i>	
Statistical Substring Reduction in Linear Time . . . . .	320
<i>Xueqiang Lü, Le Zhang, and Junfeng Hu</i>	

Detecting Sentence Boundaries in Japanese Speech Transcriptions  
Using a Morphological Analyzer . . . . . 328  
*Sachie Tajima, Hidetsugu Nanba, and Manabu Okumura*

Specification Retrieval –  
How to Find Attribute-Value Information on the Web . . . . . 338  
*Minoru Yoshida and Hiroshi Nakagawa*

## Semantic Disambiguation

Conceptual Information-Based Sense Disambiguation . . . . . 348  
*You-Jin Chung, Kyonghi Moon, and Jong-Hyeok Lee*

Influence of WSD on Cross-Language Information Retrieval . . . . . 358  
*In-Su Kang, Seung-Hoon Na, and Jong-Hyeok Lee*

Resolution of Modifier-Head Relation Gaps  
Using Automatically Extracted Metonymic Expressions . . . . . 367  
*Yoji Kiyota, Sadao Kurohashi, and Fuyuko Kido*

Word Sense Disambiguation Using Heterogeneous Language Resources . . . . 377  
*Kiyooki Shirai and Takayuki Tamagaki*

Improving Word Sense Disambiguation by Pseudo-samples . . . . . 386  
*Xiaojie Wang and Yuji Matsumoto*

## Statistical Models and Machine Learning for NLP

Long Distance Dependency in Language Modeling: An Empirical Study . . . 396  
*Jianfeng Gao and Hisami Suzuki*

Word Folding: Taking the Snapshot of Words Instead of the Whole . . . . . 406  
*Jin-Dong Kim and Jun'ichi Tsujii*

Bilingual Chunk Alignment Based on Interactional Matching  
and Probabilistic Latent Semantic Indexing . . . . . 416  
*Feifan Liu, Qianli Jin, Jun Zhao, and Bo Xu*

Learning to Filter Junk E-mail from Positive and Unlabeled Examples . . . 426  
*Karl-Michael Schneider*

A Collaborative Ability Measurement for Co-training . . . . . 436  
*Dan Shen, Jie Zhang, Jian Su, Guodong Zhou, and Chew-Lim Tan*

Flexible Margin Selection for Reranking with Full Pairwise Samples . . . . . 446  
*Libin Shen and Aravind K. Joshi*

A Comparative Study on the Use of Labeled and Unlabeled Data for Large Margin Classifiers . . . . .	456
<i>Hiroya Takamura and Manabu Okumura</i>	
Comparing Entropies Within the Chinese Language . . . . .	466
<i>Benjamin K. Tsou, Tom B.Y. Lai, and Ka-po Chow</i>	
NTPC: N-Fold Templated Piped Correction . . . . .	476
<i>Dekai Wu, Grace Ngai, and Marine Carpuat</i>	
A Three Level Cache-Based Adaptive Chinese Language Model . . . . .	487
<i>Junlin Zhang, Le Sun, Weimin Qu, Lin Du, and Yufang Sun</i>	

## Taggers, Chunkers, Shallow Parsers

Using a Smoothing Maximum Entropy Model for Chinese Nominal Entity Tagging . . . . .	493
<i>Jinying Chen, Nianwen Xue, and Martha Palmer</i>	
Deterministic Dependency Structure Analyzer for Chinese . . . . .	500
<i>Yuchang Cheng, Masayuki Asahara, and Yuji Matsumoto</i>	
High Speed Unknown Word Prediction Using Support Vector Machine for Chinese Text-to-Speech Systems . . . . .	509
<i>Juhong Ha, Yu Zheng, Byeongchang Kim, Gary Geunbae Lee, and Yoon-Suk Seong</i>	
Syntactic Analysis of Long Sentences Based on S-Clauses . . . . .	518
<i>Mi-Young Kim and Jong-Hyeok Lee</i>	
Chinese Chunk Identification Using SVMs Plus Sigmoid . . . . .	527
<i>Yongmei Tan, Tianshun Yao, Qing Chen, and Jingbo Zhu</i>	
Tagging Complex NEs with MaxEnt Models: Layered Structures Versus Extended Tagset . . . . .	537
<i>Deyi Xiong, Hongkui Yu, and Qun Liu</i>	
A Nearest-Neighbor Method for Resolving PP-Attachment Ambiguity . . . .	545
<i>Shaojun Zhao and Dekang Lin</i>	

## Text and Sentence Generation

Detection of Incorrect Case Assignments in Paraphrase Generation . . . . .	555
<i>Atsushi Fujita, Kentaro Inui, and Yuji Matsumoto</i>	
Building a Pronominalization Model by Feature Selection and Machine Learning . . . . .	566
<i>Ji-Eun Roh and Jong-Hyeok Lee</i>	

## Text Mining

Categorizing Unknown Text Segments for Information Extraction Using a Search Result Mining Approach . . . . .	576
<i>Chien-Chung Huang, Shui-Lung Chuang, and Lee-Feng Chien</i>	
Mining Table Information on the Internet . . . . .	587
<i>Sung-won Jung, Gi-deuk Han, and Hyuk-chul Kwon</i>	
Collecting Evaluative Expressions for Opinion Extraction . . . . .	596
<i>Nozomi Kobayashi, Kentaro Inui, Yuji Matsumoto, Kenji Tateishi, and Toshikazu Fukushima</i>	
A Study of Semi-discrete Matrix Decomposition for LSI in Automated Text Categorization . . . . .	606
<i>Wang Qiang, Wang XiaoLong, and Guan Yi</i>	
Systematic Construction of Hierarchical Classifier in SVM-Based Text Categorization . . . . .	616
<i>Yongwook Yoon, Changki Lee, and Gary Geunbae Lee</i>	

## Theories and Formalisms for Morphology, Syntax and Semantics

Implementing the Syntax of Japanese Numeral Classifiers . . . . .	626
<i>Emily M. Bender and Melanie Siegel</i>	
A Graph Grammar Approach to Map Between Dependency Trees and Topological Models . . . . .	636
<i>Bernd Bohnet</i>	
The Automatic Acquisition of Verb Subcategorisations and Their Impact on the Performance of an HPSG Parser . . . . .	646
<i>John Carroll and Alex C. Fang</i>	
Chinese Treebanks and Grammar Extraction . . . . .	655
<i>Keh-Jiann Chen and Yu-Ming Hsieh</i>	
FML-Based SCF Predefinition Learning for Chinese Verbs . . . . .	664
<i>Xiwu Han, Tiejun Zhao, and Muyun Yang</i>	
Deep Analysis of Modern Greek . . . . .	674
<i>Valia Kordoni and Julia Neu</i>	
Corpus-Oriented Grammar Development for Acquiring a Head-Driven Phrase Structure Grammar from the Penn Treebank . . . . .	684
<i>Yusuke Miyao, Takashi Ninomiya, and Jun'ichi Tsujii</i>	

## Word Segmentation

- Unsupervised Segmentation of Chinese Corpus Using Accessor Variety ... 694  
*Haodi Feng, Kang Chen, Chunyu Kit, and Xiaotie Deng*
- Chinese Unknown Word Identification Using Class-Based LM ..... 704  
*Guohong Fu and Kang-Kwong Luke*
- An Example-Based Study on Chinese Word Segmentation  
 Using Critical Fragments ..... 714  
*Qinan Hu, Haihua Pan, and Chunyu Kit*
- The Use of SVM for Chinese New Word Identification ..... 723  
*Hongqiao Li, Chang-Ning Huang, Jianfeng Gao, and Xiaozhong Fan*
- Chinese New Word Finding Using Character-Based Parsing Model ..... 733  
*Yao Meng, Hao Yu, and Fumihito Nishino*

## Thematic Session: Natural Language Technology in Mobile Information Retrieval and Text Processing User Interfaces

- Introduction ..... 743  
*Michael Kuehn, Mun-Kew Leong, and Kumiko Tanaka-Ishii*
- Spoken Versus Written Queries for Mobile Information Access:  
 An Experiment on Mandarin Chinese ..... 745  
*Heather Du and Fabio Crestani*
- An Interactive Proofreading System for Inappropriately Selected Words  
 on Using Predictive Text Entry ..... 755  
*Hideya Iwasaki and Kumiko Tanaka-Ishii*
- Dit4dah: Predictive Pruning for Morse Code Text Entry ..... 765  
*Kumiko Tanaka-Ishii and Ian Frank*

## Thematic Session: Text Mining in Biomedicine

- Introduction ..... 776  
*Sophia Ananiadou and Jong C. Park*
- Unsupervised Event Extraction from Biomedical Literature  
 Using Co-occurrence Information and Basic Patterns ..... 777  
*Hong-woo Chun, Young-sook Hwang, and Hae-Chang Rim*
- Annotation of Gene Products in the Literature with Gene Ontology Terms  
 Using Syntactic Dependencies ..... 787  
*Jung-jae Kim and Jong C. Park*



XVIII Table of Contents

Mining Biomedical Abstracts: What's in a Term? ..... 797  
*Goran Nenadic, Irena Spasic, and Sophia Ananiadou*

SVM-Based Biological Named Entity Recognition  
Using Minimum Edit-Distance Feature Boosted by Virtual Examples .... 807  
*Eunji Yi, Gary Geunbae Lee, Yu Song, and Soo-Jun Park*

**Author Index** ..... 815