

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 Cuřín, 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 Udapa 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
<i>Sachie Tajima, Hidetsugu Nanba, and Manabu Okumura</i>	
Specification Retrieval – How to Find Attribute-Value Information on the Web	338
<i>Minoru Yoshida and Hiroshi Nakagawa</i>	

Semantic Disambiguation

Conceptual Information-Based Sense Disambiguation	348
<i>You-Jin Chung, Kyonghi Moon, and Jong-Hyeok Lee</i>	
Influence of WSD on Cross-Language Information Retrieval	358
<i>In-Su Kang, Seung-Hoon Na, and Jong-Hyeok Lee</i>	
Resolution of Modifier-Head Relation Gaps Using Automatically Extracted Metonymic Expressions	367
<i>Yoji Kiyota, Sadao Kurohashi, and Fuyuko Kido</i>	
Word Sense Disambiguation Using Heterogeneous Language Resources	377
<i>Kiyoaki Shirai and Takayuki Tamagaki</i>	
Improving Word Sense Disambiguation by Pseudo-samples	386
<i>Xiaojie Wang and Yuji Matsumoto</i>	

Statistical Models and Machine Learning for NLP

Long Distance Dependency in Language Modeling: An Empirical Study ...	396
<i>Jianfeng Gao and Hisami Suzuki</i>	
Word Folding: Taking the Snapshot of Words Instead of the Whole	406
<i>Jin-Dong Kim and Jun'ichi Tsujii</i>	
Bilingual Chunk Alignment Based on Interactional Matching and Probabilistic Latent Semantic Indexing	416
<i>Feifan Liu, Qianli Jin, Jun Zhao, and Bo Xu</i>	
Learning to Filter Junk E-mail from Positive and Unlabeled Examples	426
<i>Karl-Michael Schneider</i>	
A Collaborative Ability Measurement for Co-training	436
<i>Dan Shen, Jie Zhang, Jian Su, Guodong Zhou, and Chew-Lim Tan</i>	
Flexible Margin Selection for Reranking with Full Pairwise Samples	446
<i>Libin Shen and Aravind K. Joshi</i>	

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

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

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

Thematic Session: Text Mining in Biomedicine

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

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

Natural Language Processing – IJCNLP 2004

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

Su, K.-Y.; Tsujii, J.; Lee, J.-H.; Kwong, O.Y. (Eds.)

2005, XVIII, 822 p., Softcover

ISBN: 978-3-540-24475-2