

Robert Dale Kam-Fai Wong Jian Su
Oi Yee Kwong (Eds.)

Natural Language Processing – IJCNLP 2005

Second International Joint Conference
Jeju Island, Korea, October 11-13, 2005
Proceedings

Table of Contents

Information Retrieval

| | |
|---|----|
| A New Method for Sentiment Classification in Text Retrieval <i>Yi Hu, Jianyong Duan, Xiaoming Chen, Bingzhen Pei, Ruzhan Lu</i> | 1 |
| Topic Tracking Based on Linguistic Features <i>Fumiyo Fukumoto, Yusuke Yamaji</i> | 10 |
| The Use of Monolingual Context Vectors for Missing Translations in Cross-Language Information Retrieval <i>Yan Qu, Gregory Grefenstette, David A. Evans</i> | 22 |
| Automatic Image Annotation Using Maximum Entropy Model <i>Wei Li, Maosong Sun</i> | 34 |

Corpus-Based Parsing

| | |
|--|----|
| Corpus-Based Analysis of Japanese Relative Clause Constructions <i>Takeshi Abekawa, Manabu Okumura</i> | 46 |
| Parsing Biomedical Literature <i>Matthew Lease, Eugene Charniak</i> | 58 |
| Parsing the Penn Chinese Treebank with Semantic Knowledge <i>Deji Xiong, Shuanglong Li, Qun Liu, Shouxun Lin, Yueliang Qian</i> | 70 |
| Using a Partially Annotated Corpus to Build a Dependency Parser for Japanese <i>Manabu Sassano</i> | 82 |

Web Mining

| | |
|---|-----|
| Entropy as an Indicator of Context Boundaries: An Experiment Using a Web Search Engine <i>Kumiko Tanaka-Ishii</i> | 93 |
| Automatic Discovery of Attribute Words from Web Documents <i>Kosuke Tokunaga, Jun'ichi Kazama, Kentaro Torisawa</i> | 106 |

| | |
|--|-----|
| Aligning Needles in a Haystack: Paraphrase Acquisition Across the Web <i>Marius Pasca, Péter Dienes</i> | 119 |
| Confirmed Knowledge Acquisition Using Mails Posted to a Mailing List <i>Yasuhiko Watanabe, Ryo Nishimura, Yoshihiro Okada</i> | 131 |

Rule-Based Parsing

| | |
|--|-----|
| Automatic Partial Parsing Rule Acquisition Using Decision Tree Induction <i>Myung-Seok Choi, Chul Su Lim, Key-Sun Choi</i> | 143 |
| Chunking Using Conditional Random Fields in Korean Texts <i>Yong-Hun Lee, Mi-Young Kim, Jong-Hyeok Lee</i> | 155 |
| High Efficiency Realization for a Wide-Coverage Unification Grammar <i>John Carroll, Stephan Oepen</i> | 165 |
| Linguistically-Motivated Grammar Extraction, Generalization and Adaptation <i>Yu-Ming Hsieh, Duen-Chi Yang, Keh-Jiann Chen</i> | 177 |

Disambiguation

| | |
|---|-----|
| PP-Attachment Disambiguation Boosted by a Gigantic Volume of Unambiguous Examples <i>Daisuke Kawahara, Sadao Kurohashi</i> | 188 |
| Adapting a Probabilistic Disambiguation Model of an HPSG Parser to a New Domain <i>Tadayoshi Hara, Yusuke Miyao, Jun'ichi Tsujii</i> | 199 |
| A Hybrid Approach to Single and Multiple PP Attachment Using WordNet <i>Akshar Bharathi, Rohini U., Vishnu P., S.M. Bendre, Rajeev Sangal</i> | 211 |
| Period Disambiguation with Maxent Model <i>Chunyu Kit, Xiaoyue Liu</i> | 223 |

Text Mining

| | |
|--|-----|
| Acquiring Synonyms from Monolingual Comparable Texts <i>Mitsuo Shimohata, Eiichiro Sumita</i> | 233 |
|--|-----|

A Method of Recognizing Entity and Relation

Xinghua Fan, Maosong Sun 245

Inversion Transduction Grammar Constraints for Mining Parallel Sentences from Quasi-Comparable Corpora

Dekai Wu, Pascale Fung 257

Automatic Term Extraction Based on Perplexity of Compound Words

Minoru Yoshida, Hiroshi Nakagawa 269

Document Analysis

Document Clustering with Grouping and Chaining Algorithms

Yllias Chali, Soufiane Noureddine 280

Using Multiple Discriminant Analysis Approach for Linear Text Segmentation

Zhu Jingbo, Ye Na, Chang Xinzhi, Chen Wenliang, Benjamin K. Tsou 292

Classifying Chinese Texts in Two Steps

Xinghua Fan, Maosong Sun, Key-sun Choi, Qin Zhang 302

Assigning Polarity Scores to Reviews Using Machine Learning Techniques

Daisuke Okanohara, Jun'ichi Tsujii 314

Ontology and Thesaurus

Analogy as Functional Recategorization: Abstraction with HowNet Semantics

Tony Veale 326

PLSI Utilization for Automatic Thesaurus Construction

Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama 334

Analysis of an Iterative Algorithm for Term-Based Ontology Alignment

Shisanu Tongchim, Canasai Kruengkrai, Virach Sornlertlamvanich, Prapass Srichaivattana, Hitoshi Isahara 346

Finding Taxonomical Relation from an MRD for Thesaurus Extension

SeonHwa Choi, HyukRo Park 357

Relation Extraction

| | |
|--|-----|
| Relation Extraction Using Support Vector Machine <i>Gumwon Hong</i> | 366 |
| Discovering Relations Between Named Entities from a Large Raw Corpus Using Tree Similarity-Based Clustering <i>Min Zhang, Jian Su, Danmei Wang, Guodong Zhou, Chew Lim Tan</i> | 378 |
| Automatic Relation Extraction with Model Order Selection and Discriminative Label Identification <i>Chen Jinxiu, Ji Donghong, Tan Chew Lim, Niu Zhengyu</i> | 390 |
| Mining Inter-Entity Semantic Relations Using Improved Transductive Learning <i>Zhu Zhang</i> | 402 |

Text Classification

| | |
|---|-----|
| A Preliminary Work on Classifying Time Granularities of Temporal Questions <i>Wei Li, Wenjie Li, Qin Lu, Kam-Fai Wong</i> | 414 |
| Classification of Multiple-Sentence Questions <i>Akihiro Tamura, Hiroya Takamura, Manabu Okumura</i> | 426 |

Transliteration

| | |
|--|-----|
| A Rule Based Syllabification Algorithm for Sinhala <i>Ruvan Weerasinghe, Asanka Wasala, Kumudu Gamage</i> | 438 |
| An Ensemble of Grapheme and Phoneme for Machine Transliteration <i>Jong-Hoon Oh, Key-Sun Choi</i> | 450 |

Machine Translation – I

| | |
|--|-----|
| Improving Statistical Word Alignment with Ensemble Methods <i>Wu Hua, Wang Haifeng</i> | 462 |
| Empirical Study of Utilizing Morph-Syntactic Information in SMT <i>Young-Sook Hwang, Taro Watanabe, Yutaka Sasaki</i> | 474 |

Question Answering

Instance-Based Generation for Interactive Restricted Domain Question Answering Systems

Matthias Denecke, Hajime Tsukada 486

Answering Definition Questions Using Web Knowledge Bases

Zhushuo Zhang, Yaqian Zhou, Xuanjing Huang, Lide Wu 498

Exploring Syntactic Relation Patterns for Question Answering

Dan Shen, Geert-Jan M. Kruijff, Dietrich Klakow 507

Web-Based Unsupervised Learning for Query Formulation in Question Answering

*Yi-Chia Wang, Jian-Cheng Wu, Tyne Liang,
Jason S. Chang* 519

Morphological Analysis

A Chunking Strategy Towards Unknown Word Detection in Chinese Word Segmentation

Zhou GuoDong 530

A Lexicon-Constrained Character Model for Chinese Morphological Analysis

Yao Meng, Hao Yu, Fumihito Nishino 542

Relative Compositionality of Multi-word Expressions: A Study of Verb-Noun (V-N) Collocations

Sriram Venkatapathy, Aravind K. Joshi 553

Automatic Extraction of Fixed Multiword Expressions

Campbell Hore, Masayuki Asahara, Yūji Matsumoto 565

Machine Translation – II

Phrase-Based Statistical Machine Translation: A Level of Detail Approach

Hendra Setiawan, Haizhou Li, Min Zhang, Beng Chin Ooi 576

Why Is Zero Marking Important in Korean?

Sun-Hee Lee, Donna K. Byron, Seok Bae Jang 588

| | |
|---|-----|
| A Phrase-Based Context-Dependent Joint Probability Model for Named Entity Translation <i>Min Zhang, Haizhou Li, Jian Su, Hendra Setiawan</i> | 600 |
| Machine Translation Based on Constraint-Based Synchronous Grammar <i>Fai Wong, Dong-Cheng Hu, Yu-Hang Mao, Ming-Chui Dong, Yi-Ping Li</i> | 612 |
| Text Summarization | |
| A Machine Learning Approach to Sentence Ordering for Multidocument Summarization and Its Evaluation <i>Danushka Bollegala, Naoaki Okazaki, Mitsuru Ishizuka</i> | 624 |
| Significant Sentence Extraction by Euclidean Distance Based on Singular Value Decomposition <i>Changbeom Lee, Hyukro Park, Cheolyoung Ock</i> | 636 |
| Named Entity Recognition | |
| Two-Phase Biomedical Named Entity Recognition Using a Hybrid Method <i>Seonho Kim, Juntae Yoon, Kyung-Mi Park, Hae-Chang Rim</i> | 646 |
| Heuristic Methods for Reducing Errors of Geographic Named Entities Learned by Bootstrapping <i>Seungwoo Lee, Gary Geunbae Lee</i> | 658 |
| Linguistic Resources and Tools | |
| Building a Japanese-Chinese Dictionary Using Kanji/Hanzi Conversion <i>Chooi-Ling Goh, Masayuki Asahara, Yuji Matsumoto</i> | 670 |
| Automatic Acquisition of Basic Katakana Lexicon from a Given Corpus <i>Toshiaki Nakazawa, Daisuke Kawahara, Sadao Kurohashi</i> | 682 |
| CTEMP: A Chinese Temporal Parser for Extracting and Normalizing Temporal Information <i>Wu Mingli, Li Wenjie, Lu Qin, Li Baoli</i> | 694 |
| French-English Terminology Extraction from Comparable Corpora <i>Béatrice Daille, Emmanuel Morin</i> | 707 |

Discourse Analysis

A Twin-Candidate Model of Coreference Resolution with Non-Anaphor Identification Capability

Xiaofeng Yang, Jian Su, Chew Lim Tan 719

Improving Korean Speech Acts Analysis by Using Shrinkage and Discourse Stack

Kyungsun Kim, Youngjoong Ko, Jungyun Seo 731

Anaphora Resolution for Biomedical Literature by Exploiting Multiple Resources

Tyne Liang, Yu-Hsiang Lin 742

Automatic Slide Generation Based on Discourse Structure Analysis

Tomohide Shibata, Sadao Kurohashi 754

Semantic Analysis – I

Using the Structure of a Conceptual Network in Computing Semantic Relatedness

Iryna Gurevych 767

Semantic Role Labelling of Prepositional Phrases

Patrick Ye, Timothy Baldwin 779

Global Path-Based Refinement of Noisy Graphs Applied to Verb Semantics

Timothy Chklovski, Patrick Pantel 792

Semantic Role Tagging for Chinese at the Lexical Level

Oi Yee Kwong, Benjamin K. Tsou 804

NLP Applications

Detecting Article Errors Based on the Mass Count Distinction

Ryo Nagata, Takahiro Wakana, Fumito Masui, Atsuo Kawai, Naoki Izu 815

Principles of Non-stationary Hidden Markov Model and Its Applications to Sequence Labeling Task

Xiao JingHui, Liu BingQuan, Wang XiaoLong 827

| | |
|--|-----|
| Integrating Punctuation Rules and Naïve Bayesian Model for Chinese Creation Title Recognition <i>Conrad Chen, Hsin-Hsi Chen</i> | 838 |
| A Connectionist Model of Anticipation in Visual Worlds <i>Marshall R. Mayberry, III, Matthew W. Crocker, Pia Knoeferle</i> | 849 |
| Tagging | |
| Automatically Inducing a Part-of-Speech Tagger by Projecting from Multiple Source Languages Across Aligned Corpora <i>Victoria Fossum, Steven Abney</i> | 862 |
| The Verbal Entries and Their Description in a Grammatical Information-Dictionary of Contemporary Tibetan <i>Jiang Di, Long Congjun, Zhang Jichuan</i> | 874 |
| Tense Tagging for Verbs in Cross-Lingual Context: A Case Study <i>Yang Ye, Zhu Zhang</i> | 885 |
| Regularisation Techniques for Conditional Random Fields: Parameterised Versus Parameter-Free <i>Andrew Smith, Miles Osborne</i> | 896 |
| Semantic Analysis – II | |
| Exploiting Lexical Conceptual Structure for Paraphrase Generation <i>Atsushi Fujita, Kentaro Inui, Yuji Matsumoto</i> | 908 |
| Word Sense Disambiguation by Relative Selection <i>Hee-Cheol Seo, Hae-Chang Rim, Myung-Gil Jang</i> | 920 |
| Towards Robust High Performance Word Sense Disambiguation of English Verbs Using Rich Linguistic Features <i>Jinying Chen, Martha Palmer</i> | 933 |
| Automatic Interpretation of Noun Compounds Using WordNet Similarity <i>Su Nam Kim, Timothy Baldwin</i> | 945 |
| Language Models | |
| An Empirical Study on Language Model Adaptation Using a Metric of Domain Similarity <i>Wei Yuan, Jianfeng Gao, Hisami Suzuki</i> | 957 |

| | |
|---|----------|
| A Comparative Study of Language Models for Book and Author Recognition | |
| <i>Özlem Uzuner, Boris Katz</i> | 969 |
| Spoken Language | |
| Lexical Choice via Topic Adaptation for Paraphrasing Written Language to Spoken Language | |
| <i>Nobuhiro Kaji, Sadao Kurohashi</i> | 981 |
| A Case-Based Reasoning Approach for Speech Corpus Generation | |
| <i>Yandong Fan, Elizabeth Kendall</i> | 993 |
| Terminology Mining | |
| Web-Based Terminology Translation Mining | |
| <i>Gaolin Fang, Hao Yu, Fumihito Nishino</i> | 1004 |
| Extracting Terminologically Relevant Collocations in the Translation of Chinese Monograph | |
| <i>Byeong-Kwu Kang, Bao-Bao Chang, Yi-Rong Chen, Shi-Wen Yu</i> | 1017 |
| Author Index | 1029 |