
IJCNLP 2008

**Third International Joint Conference
on
Natural Language Processing**

Proceedings of the Conference

Organizer

Asian Federation of Natural Language Processing

Local Host

International Institute of Information Technology, India

January 7-12 2008
Hyderabad, India

Hosts/Organizers

Asian Federation of Natural Language Processing (AFNLP)

International Institute of Information Technology, Hyderabad, India (IIIT-H)

Natural Language Processing Association of India (NLP AI)

Supporters

Yahoo! Research & Development, India

Department of Information Technology, MCIT, Government of India

Information & Communication Technology, Government of Andhra Pradesh

Microsoft Research, India

Tata Consultancy Services Limited

Centre for Development of Advanced Computing

Indian School of Business

Satyam Computer Service Limited

Google India

IBM India

Defense Research and Development Organization

Council for Scientific and Industrial Research

Preface: Conference Chair

Dear colleagues,

Welcome to the 2008 International Joint Conference on Natural Language Processing (IJCNLP-08). This is the third biennial conference organized by the Asian Federation of Natural Language Processing (AFNLP), which was founded in 2004 to promote research and development efforts in the field of computational processing of natural languages of importance to the Asian region, without regard to differences in language, race, religious belief or political stand. The first IJCNLP was held to celebrate the inauguration of AFNLP on the beautiful Hainan Island in China (March 22-24, 2004), and the second on the fantastic Jeju Island in Korea (October 10-13, 2005). Following the continuing success of the previous two conferences, the third conference is held in yet another exotic and multicultural city of Hyderabad in India in January 7-12, 2008.

On behalf of the Conference Committees, I would like to welcome all researchers and scholars who are working in all areas of Natural Language Processing (NLP) around the world and who in particular have keen interest in Asian language processing. As the world proceeds quickly into the Information Age, we face both successes and challenges in creating a global information society, and it is well recognized nowadays that Natural Language Processing provides the key to the Information Age and to solving many of these challenges, like breaking language barrier and overcoming information flood. Over the last decades, a remarkable progress has been made in NLP research and development. However, there has been a pervasive feeling that the progress of NLP for Asian languages has not been commensurate with that for Western languages. Recently the importance of Asian languages has been steadily growing as Asia becomes the dominant region of the world, economically, politically and culturally. In this context, this conference provides a forum for engineers and scientists to present and exchange their latest research findings in all aspects of NLP and thus to promote research and development activities for Asian language processing. This is the major motivation of IJCNLP.

I would like to express my sincere appreciation to the authors of invited and contributed papers and to all conference participants for their active participations. I also wish to express my heartfelt gratitude and thanks to the Committee Members, particularly the Organizing Co-chairs Rajeev Sangal and Raji Bagga, the Program Co-chairs Yuji Matsumoto and Ann Copestake, the Publication Chair Jing-Shin Chang, and all the other Committee Chairs for their tremendous efforts and substantial contributions to the conference. I feel honored and blessed to be part of this conference as the Conference Chair working with such wonderful team. With our team efforts, I am confident that this conference will be even more successful than the previous. Finally, I hope that you will participate actively in all sessions and events to maximize the benefits from them, and I also wish all participants a very fruitful and enjoyable time during the conference in Hyderabad.

Jong-Hyeok Lee
Conference Chair

Preface: Program Committee Co-Chairs

This volume contains the papers accepted for presentation at the third International Joint Conference on Natural Language Processing (IJCNLP-2008). IJCNLP is held approximately every two years as the flagship conference of the AFNLP (Asian Federation of Natural Language Processing). This year's conference, which follows the success of IJCNLP-2005 on Jeju Island in Korea, is in the city which is such a beautiful mixture of ancient civilization and modern industry: Hyderabad, India.

On behalf of the Program Committee, we are pleased to present this volume, which includes the accepted papers for oral and poster presentations at the conference. We received 266 submissions from 28 different regions all over the world; 74% from Asia, 15% from North America, 9% from Europe, 3% from Australia, and 0.4% from Africa.

The paper selection was not easy with this large number of submission but with the devoted work of our 13 area chairs and 268 PC members, we were able to select very high quality papers. 75 papers (27.8%) were accepted for oral presentation and 62 papers (23.3%) were accepted for poster presentation. After a few withdrawals, this volume contains 74 oral papers and 60 poster papers.

We are also very grateful to Professor Aravind Joshi (University of Pennsylvania), Professor Hyopil Shin (Seoul National University) and Dr S.H. Srinivasan (Yahoo!) for accepting to give keynote and invited talks, which surely make the conference more attractive.

Organizing and hosting this size of international conference requires lots of help and effort from many people. We would like to send our greatest thanks to the Honorary Conference Chair, Professor Aravind Joshi and the Conference Chair, Professor Jong-Hyoek Lee for their continuous support and timely guidance. We would also thank the Local Organizing Co-Chairs Professor Rajeev Sangal and Dr Raji Bagga for their support, advice and responses to our numerous requests. Special thanks are due to the Publication Chair Professor Jing-Shin Chang. Without his very detailed format checking and efficient compilation, this volume of proceedings would not come out in the current form. We would like to express our highest gratitude to all the other Committee Chairs. Working with such a wonderful group of people has been great fun. Last but not least, we would like to thank all the people who submitted their papers and all the people who attend this conference.

Welcome to IJCNLP-2008. We hope you enjoy this conference as much as we do.

Ann Copestake and Yuji Matsumoto
Program Committee Co-Chairs

Keynote Speech:

PENN Discourse Treebank: Complexity of Dependencies at the Discourse Level and at the Sentence Level

Aravind K. Joshi

Department of Computer and Information Science and
Institute for Research in Cognitive Science
University of Pennsylvania, Philadelphia, PA, USA

ABSTRACT

First, I will describe the Penn Discourse Treebank (PDTB)*, a corpus in which we annotate the discourse connectives (explicit and implicit) and their arguments, together with "attributions" of the arguments and the relations denoted by the connectives, and also the senses of the connectives. I will then discuss some issues concerning the complexity of dependencies in terms of the elements that bear the dependency relations, the graph theoretic properties of these dependencies such as nested and crossed dependencies, dependencies with shared arguments, and finally, the attributions and their relationship to the dependencies, among others. We will compare these dependencies with those at the sentence level and then discuss some aspects that relate to the transition from the sentence level to the level of "immediate discourse" and propose some conjectures.

* This 1 million-word corpus is the same as the WSJ corpus used by the Penn Treebank (PTB) for syntactic annotation and by Propbank for predicate-argument annotation. PDTB 2.0 will be released by the Linguistic Data Consortium (LDC) in early February 2008.

Members of the PDTB project: Nikhil Dinesh, Aravind K. Joshi, Alan Lee, Eleni Miltsakaki, Rashmi Prasad, and Bonnie Webber (University of Edinburgh).

Invited Talk:

The 21st Sejong Project: with a Focus on Building of the SELK (Sejong Electronic Lexicon of Korean) and the KNC (Korean National Corpus)

Hyopil Shin

Dept. of Linguistics, Seoul National University
School of Computer Engineering, Seoul National University

ABSTRACT

The 21st Sejong Project started in 1998 with a 10-year plan. The project was funded by the Ministry of Culture and Tourism of the Korean government. The goal of the project was to promote technological expertise in Korean language research and technology. The project consists of 8 sub-projects ranging from construction of Korean language resources to management and distribution of outputs from the work. The core part of the project is to compile an electronic lexical dictionary and to build a large-scale Korean corpus.

The SELK focuses on an exhaustive representation of Korean linguistic knowledge by harmonizing linguistic validity, psychological reality, and computational efficiency. The SELK is composed of various sub-dictionaries corresponding to the parts-of-speech-based word categories such as nouns, verbs, adverbs etc. The lexicon shows a considerable differentiation from other paperback or machine-readable dictionaries in Korean in its precise and comprehensive representation.

The KNC project has two sub-divisions, one for a general corpus and the other for a special corpus. The general corpus division collected a wide range of unconstrained materials and endeavored at annotating the data with parts-of-speech, syntactic, and semantic tags. The special data division, on the other hand, constructed Korean-English and Korean-Japanese corpora, a historical corpus, and a corpus used by North Koreans and overseas Koreans.

The SELK and the KNC are beginning to serve as important research tools for investigators in natural language processing as well as in theoretical linguistics. Annotated corpora and well-established electronic dictionaries promise to be valuable for enterprises such as the construction of statistical models for the grammar of written and spoken Korean, the development of software for Korean language processing, and even the publication of the paperback Korean dictionaries.

In this speech, I will introduce the 21st Sejong Project and review my experience with constructing one such large language resource - the SELK, consisting of about 600,000 lexical entries, and the KNC, consisting of about 500 million word collections. Considering the size and time needed to develop it, this project deserves great attention. We, however, also experienced a lot of difficulties through trial and error, inevitably originating from such a long work period and the large scale of the work. We hope sharing such experiences will help researchers with the same interests, to break through the obstacles and to avoid mistakes we have made for a decade.

Invited Talk:

Language Processing for the Evolving Web

Srinivasan Sengamedu

Yahoo!, Bangalore

ABSTRACT

World Wide Web brings several new dimensions to language processing - social, multimodal, structural, etc. The social dimension arises from the tagging phenomenon, multimodal from the coexistence of images and videos with text in web documents, and structure from rich formatting of web pages. While the massive amounts of data available has made new approaches to translation, summarization, and extraction possible, next generation applications like semantic search require radically new theoretical ideas. The talk will outline the phenomena, summarize recent achievements, and pose the new challenges.

Conference Committee

Honorary Conference Chair

Prof. Aravind Joshi, University of Pennsylvania

Conference Chair

Prof. Jong-Hyeok Lee, POSTECH, Korea

Local Organizing Co-chairs

Prof. Rajeev Sangal, International Institute of Information Technology, India

Dr. Raji Bagga, International Institute of Information Technology, India

LOC Steering/Core Committee Members

Dr. A Kumaran, Microsoft Research India

Dr. Pushpak Bhattacharya, IIT Bombay, India

Dr. M Sasikumar, CDAC, Mumbai, India

Dr. Dipti Misra Sharma, International Institute of Information Technology, India

Dr. Vasudeva Varma, International Institute of Information Technology, India

Dr. KS Rajan, International Institute of Information Technology, India

Dr. Anoop M Namboodri, International Institute of Information Technology, India

Mr. KS Vijaya Sekhar, IJCNLP-08 LOC Secretariat

Program Committee Co-chairs

Prof. Yuji Matsumoto, NAIST, Japan

Prof. Ann Copestake, University of Cambridge, UK

Publication Committee Chair

Prof. Jing-Shin Chang, National Chi-Nan University, Taiwan

Publicity Committee Co-chairs

Prof. Satoshi Sato, Nagoya University, Japan

Prof. Jian-Yun Nie, Univ. of Montreal, Canada

Financial Committee Co-chairs

Prof. Kam-Fai Wong, Chinese Univ. of HK, Hong Kong

Dr. A. Kumaran, Microsoft Research India

Tutorial Committee Co-chairs

Prof. Kemal Oflazer, Sabanci University, Turkey

Prof. Key-Sun Choi, KAIST, Korea

Workshops Committee Co-chairs

Dr. Haizhou Li, Institute of Infocomm Research, Singapore

Dr. Timothy Baldwin, University of Melbourne, Australia

Exhibition/Demo Committee Co-chairs

Prof. Pushpak Bhattacharya, Indian Institute of Technology Bombay, India

Prof. Fei Xia, Univ. of Washington, USA

Program Committee

Program Committee Chairs:

Ann Copestake, University of Cambridge
(Ann.Copestake@cl.cam.ac.uk)

Yuji Matsumoto, Nara Institute of Science and Technology
(matsu@is.naist.jp)

Areas and Area Chairs:

Information Retrieval

Hang Li, Microsoft China

Parsing/Grammatical Formalisms

Beth Ann Hockey, UCSC

Chunking/Shallow Parsing

Srinivas Bangalore, AT&T Research

Statistical Models/Machine Learning for NLP

Robert Malouf, San Diego State University

Machine Translation

Philipp Koehn, University of Edinburgh

Word Segmentation/POS Tagging

Yuji Matsumoto, Nara Institute of Science and Technology

Semantic Processing/Lexical Semantics

Patrick St Dizier, IRIT

Ontologies and Linguistic Resources

Nancy Ide, Vassar College

Paraphrasing/Entailment/Generation

Kentaro Inui, NAIST

Discourse

Alistair Knott, University of Otago

QA/Text Summarization

Sanda Harabagiu, University of Texas at Dallas

Text Mining/Information Extraction

James Curran, University of Sydney

Spoken Language Processing

Gary Geunbae Lee, POSTECH

Asian Language Processing and Linguistics Issues in NLP

Chu-Ren Huang, Academia Sinica

Program Committee Members:

Eugene Agichtein, Yaser Al-Onaizan, James Allen, Pascal Amsili, Alina Andreevskaia, Kenji Araki, Nick Asher

Collin Baker, Jason Baldridge, Timothy Baldwin, Sivaji Bandyopadhyay, Srinivas angalore, Marco Baroni, Stephen Beale, Tilman Becker, Susham Bendre, Pushpak Bhattacharyya, Zhao Bing, Sasha Blair-Goldensohn, Francis Bond, Kalina ontcheva, Johan Bos, Pierrette Bouillon, Antonio Branco, Thorsten Brants, Paul Buitelaar, Razvan Bunescu, Harry Bunt, Miriam Butt, Ekaterina Buyko, Bill Byrne

Aoife Cahill, Chris Callison-Burch, Nicoletta Calzolari, Joyce Chai, Hsin-Hsi Chen, Keh-Jiann Chen, David Chiang, Massimiliano Ciaramita, Philipp Cimiano, tephen Clark, Simon Clematide, Trevor Cohn, Nigel Collier, John Conroy, Nick Craswell, Dan Cristea, Andras Csomai, Silviu Cucerzan, Hang Cui

Walter Daelemans, Robert Dale, Hal Daume, Thierry Declerck, Fernando Diaz, Hakkani-Tur Dilek, Pavel Dmitriev, Bill Dolan, Christy Doran, Bonnie Dorr, Mark Dras, Markus Dreyer

Cecile Fabre, Hui Fang, Marcello Federico, Christiane Fellbaum, Dan Flickinger, J.Foster, Alex Fraser, Atsushi Fujita

Bin Gao, Claire Gardent, Albert Gatt, Tanja Gaustad, Niyu Ge, Dan Gildea, Jade Goldstein-Stewart, Ralph Grishman, Claire Grover, Narendra Gupta

Patrick Haffner, Jan Hajic, Keith Hall, Sanda Harabagiu, Samer Hassan, Toby Hawker, Mary Hearne, John Henderson, Andrew Hickl, Deidre Hogan, Tracy Holloway King, Shu-Kai Hsieh, Sarmad Hussain

Kentaro Inui, Hitoshi Isahara, Masato Ishizaki

Donghong Ji

Laura Kallmeyer, Kyoko Kanzaki, Tatsuya Kawahara, Asanee Kawtrakul, Junichi Kazama, Andrew Kehler, Bernd Kiefer, Adam Killgariff, Byeongchang Kim, Jin Dong Kim, Atanis Kiryakov, Manfred Klenner, Kevin Knight, Anna Korhonen, Emiel Kraemer, Shankar Kumar, Sadoa Kurohashi

Philippe Langlais, Mirella Lapata, Alon Lavie, Kiyong Lee, Alessandro Lenci, Yves Lepage, Baoli Li, Haizhou Li, Wenjie Li, Chin-Yew Lin, Tie-Yan Liu, Adam Lopez

Ryan MacDonald, Bernardo Magnini, Steve Maiorano, Robert Malouf, Gideon Mann, Alda Mari, Katja Markert, Carlos Martin-Vide, Kathleen McCoy, Ryan McDonald, Marge McShane, Michael McTear, Arul Menezes, Helen Meng, Wolfgang Minker, Vibhu Mittal, Yusuke Miyao, Paola Monachesi, Bob Moore, Tony Mullen

Program Committee Members (cont.) :

Sobha Nair, Hiroshi Nakagawa, Seiichi Nakagawa, Satoshi Nakamura, Shri Narayanan, Srinu Narayanan, Vivi Nastase, Tetsuya Nasukawa, Roberto Navigli, Vincent Ng, Orace Ngai, Thi Minh Huyen Nguyen, Jianyun Nie, Joakim Nivre, Yongkyoon No, Tadashi Nomoto, Eric Nyberg

Franz Och, Stephen Oepen, Kemal Oflazer, Miles Osborne, Lilja Ovrelid

Martha Palmer, Shimei Pan, Antonio Pareja-Lora, Jong Park, Jon Patrick, Wim Peters, Manfred Pinkal, Paul Piwek, Thierry Poibeau, David Powers, Kishore Prahallad, Rashmi Prasad

Chris Quirk

Allan Ramsay, Lance Ramshaw, Manny Rayner, Christian Retore, Brian Roark, Horacio Rodri'guez, Antti-Veikko Rosti, Rachel E. O. Roxas, Thomas Russ

Kenji Sagae, Horacio Saggion, Patrick Saint-Dizier, Ted Sanders, Rajeev Sangal, Anoop Sarkar, Sudeshna Sarkar, Holger Schwenk, Donia Scott, Satoshi Sekine, Burr Settles, Vijay Shanker, Dipiti Misra Sharma, Libin Shen, Kiyooki Shirai, Candy Sidner, Khalil Simaan, Michel Simard, Navjyoti Singh, Noah A Smith, David Smith, Virach Sornlertlamvanich, Wilbert Spooren, Caroline Sporleder, Manfred Stede, Mark Stevenson, Kristina Striegnitz, Michael Strube, Craig Struble, Jian Su, Mihai Sudreanu, Eiichiro Sumita, Yoshimi Suzuki, Hisami Suzuki

Jie Tang, Simone Teufel, Thanaruk Theeramunkong, Jo"rg Tiedemann, Christoph Tillmann, Ivan Titov, Takenobu Tokunaga, Shisanu Tongchim, Kentaro Torisawa, Kristina Toutanova, David Traum, Shu-Chuan Tseng, Yoshimasa Tsuruoka

Nicola Ueffing

Ielka van der Sluis, Menno van Zaanen, Lucy Vanderwende, Vasudeva Varma, Sriram Venkatapathy, Ashish Venugopal, Jette Viethen, Andreas Vlachos, Stephan Vogel

Marilyn Walker, Hsin-min Wang, Taro Watanabe, Bonnie Webber, Ralph Weischedel, Casey Whitelaw, Yuk Wah Wong, Yunfang Wu, Dekai Wu

Peng Xu, Jun Xu, Nianwen Xue

Naoki Yoshinaga, Yong Yu

Annie Zaenen, Richard Zens, Jun Zhao, Ming Zhou, GuoDong Zhou, Jing-bo Zhu, Michael Zock

Table of Contents

Volume I

Preface : Conference Chair.....i

Preface : Program Committee Co-Chairs.....ii

Keynote Speech and Invited Talks

PENN Discourse Treebank: Complexity of Dependencies at the Discourse Level and at the Sentence Level

Aravind K. Joshi.....iii

The 21st Sejong Project: with a Focus on Building of the SELK (Sejong Electronic Lexicon of Korean) and the KNC (Korean National Corpus)

Hyopil Shin.....iv

Language Processing for the Evolving Web

Srinivasan Sengamedu.....v

Word Segmentation/POS Tagging

A Lemmatization Method for Modern Mongolian and its Application to Information Retrieval

Badam-Osor Khaltar and Atsushi Fujii.....1

An Empirical Comparison of Goodness Measures for Unsupervised Chinese Word Segmentation with a Unified Framework

Hai Zhao and Chunyu Kit.....9

A Hybrid Approach to the Induction of Underlying Morphology

Michael Tepper and FeiXia.....17

Text Mining/Information Extraction (1)

Context-Sensitive Convolution Tree Kernel for Pronoun Resolution

GuoDong Zhou, Fang Kong and QiaoMing Zhu.....25

Semi-Supervised Learning for Relation Extraction

GuoDong Zhou, JunHui Li, LongHua Qian and QiaoMing Zhu.....32

<i>Story Link Detection based on Dynamic Information Extending</i> Xiaoyan Zhang, Ting Wang and Huowang Chen.....	40
----------------------------------------------------------------------------------------------------------------------	----

Asian Language Processing and Linguistics Issues

<i>Orthographic Disambiguation Incorporating Transliterated Probability</i> Eiji Aramaki, Takeshi Imai, Kengo Miyo and Kazuhiko Ohe.....	48
---------------------------------------------------------------------------------------------------------------------------------------------	----

<i>Name Origin Recognition Using Maximum Entropy Model and Diverse Features</i> Min Zhang, Chengjie Sun, Haizhou Li, AiTi Aw, Chew Lim Tan and Xiaolong Wang.....	56
----------------------------------------------------------------------------------------------------------------------------------------------------------------------	----

<i>A More Discerning and Adaptable Multilingual Transliteration Mechanism for Indian Languages</i> Harshit Surana and Anil Kumar Singh.....	64
------------------------------------------------------------------------------------------------------------------------------------------------	----

Parsing/Grammar

<i>UCSG: A Wide Coverage Shallow Parsing System</i> G. Bharadwaja Kumar and Kavi Narayana Murthy.....	72
----------------------------------------------------------------------------------------------------------	----

<i>Memory-Inductive Categorical Grammar: An Approach to Gap Resolution in Analytic-Language Translation</i> Prachya Boonkwan and Thepchai Supnithi.....	80
------------------------------------------------------------------------------------------------------------------------------------------------------------	----

<i>Dependency Parsing with Short Dependency Relations in Unlabeled Data</i> Wenliang Chen, Daisuke Kawahara, Kiyotaka Uchimoto, Yujie Zhang and Hitoshi Isahara.....	88
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----

Text Mining/Information Extraction (2)

<i>Effective Compositional Model for Lexical Alignment</i> Beatrice Daille and Emmanuel Morin.....	95
-------------------------------------------------------------------------------------------------------	----

<i>Determining the Unithood of Word Sequences Using a Probabilistic Approach</i> Wilson Wong, Wei Liu and Mohammed Bennamoun.....	103
--------------------------------------------------------------------------------------------------------------------------------------	-----

<i>Lexical Chains as Document Features</i> Dinakar Jayarajan, Dipti Deodhare and Balaraman Ravindran.....	111
--------------------------------------------------------------------------------------------------------------	-----

Summarization

<i>Entity-driven Rewrite for Multi-document Summarization</i> Ani Nenkova.....	118
-----------------------------------------------------------------------------------	-----

<i>A New Approach to Automatic Document Summarization</i> Xiaofeng Wu and Chengqing Zong.....	126
<i>Generic Text Summarization Using Probabilistic Latent Semantic Indexing</i> Harendra Bhandari, Masashi Shimbo, Takahiko Ito and Yuji Matsumoto.....	133
Multiple Document Processing	
<i>Identifying Cross-Document Relations between Sentences</i> Yasunari Miyabe, Hiroya Takamura and Manabu Okumura.....	141
<i>Experiments on Semantic-based Clustering for Cross-document Coreference</i> Horacio Saggion.....	149
<i>Modeling Context in Scenario Template Creation</i> Long Qiu, Min-Yen Kan and Tat-Seng Chua.....	157
<i>Cross Language Text Categorization Using a Bilingual Lexicon</i> Ke Wu, Xiaolin Wang and Bao-Liang Lu.....	165
Information Retrieval	
<i>Identify Temporal Websites Based on User Behavior Analysis</i> Yong Wang, Yiqun Liu, Min Zhang, Shaoping Ma and Liyun Ru.....	173
<i>A Comparative Study for Query Translation using Linear Combination and Confidence Measure</i> Youssef Kadri and Jian-Yun Nie.....	181
<i>TSUBAKI: An Open Search Engine Infrastructure for Developing New Information Access Methodology</i> Keiji Shinzato, Tomohide Shibata, Daisuke Kawahara, Chikara Hashimoto and Sadao Kurohashi.....	189
<i>A Study on Effectiveness of Syntactic Relationship in Dependence Retrieval Model</i> Fan Ding and Bin Wang.....	197
Spoken Language Processing	
<i>Automatic Estimation of Word Significance oriented for Speech-based Information Retrieval</i> Takashi Shichiri, Hiroaki Nanjo and Takehiko Yoshimi.....	204

<i>Rapid Prototyping of Robust Language Understanding Modules for Spoken Dialogue Systems</i> Yuichiro Fukubayashi, Kazunori Komatani, Mikio Nakano, Kotaro Funakoshi, Hiroshi Tsujino, Tetsuya Ogata and Hiroshi G. Okuno.....	210
<i>Automatic Prosodic Labeling with Conditional Random Fields and Rich Acoustic Features</i> Gina-Anne Levow.....	217
Machine Translation (1)	
<i>Chinese Unknown Word Translation by Subword Re-segmentation</i> Ruiqiang Zhang and Eiichiro Sumita.....	225
<i>Hypothesis Selection in Machine Transliteration: A Web Mining Approach</i> Jong-Hoon Oh and Hitoshi Isahara.....	233
<i>What Prompts Translators to Modify Draft Translations? An Analysis of Basic Modification Patterns for Use in the Automatic Notification of Awkwardly Translated Text</i> Takeshi Abekawa and Kyo Kageura.....	241
<i>Improving Word Alignment by Adjusting Chinese Word Segmentation</i> Ming-Hong Bai, Keh-Jiann Chen and Jason S. Chang.....	249
Text Mining/Information Extraction (3)	
<i>The Telling Tail: Signals of Success in Electronic Negotiation Texts</i> Marina Sokolova, Vivi Nastase and Stan Szpakowicz.....	257
<i>Automatic Extraction of Briefing Templates</i> Dipanjan Das, Mohit Kumar and Alexander I. Rudnicky.....	265
<i>Mining the Web for Relations between Digital Devices using a Probabilistic Maximum Margin Model</i> Oksana Yakhnenko and Barbara Rosario.....	273
<i>Learning Patterns from the Web to Translate Named Entities for Cross Language Information Retrieval</i> Yu-Chun Wang, Richard Tzong-Han Tsai and Wen-Lian Hsu.....	281

Emotion/Sentiment Analysis

<i>Bootstrapping Both Product Features and Opinion Words from Chinese Customer Reviews with Cross-Inducing</i>	
Bo Wang and Houfeng Wang.....	289
<i>Learning to Shift the Polarity of Words for Sentiment Classification</i>	
Daisuke Ikeda, Hiroya Takamura, Lev-Arie Ratinov and Manabu Okumura.....	296
<i>Unsupervised Classification of Sentiment and Objectivity in Chinese Text</i>	
Taras Zagibalov and John Carroll.....	304
<i>Using Roget's Thesaurus for Fine-grained Emotion Recognition</i>	
Saima Aman and Stan Szpakowicz.....	312

Machine Translation (2)

<i>Heterogeneous Automatic MT Evaluation Through Non-Parametric Metric Combinations</i>	
Jesús Giménez and Lluís Márquez.....	319
<i>Paraphrasing depending on Bilingual Context Toward Generalization of Translation Knowledge</i>	
Young-Sook Hwang, Young-Kil Kim and Sangkyu Park.....	327

Named Entity Recognition

<i>A Framework Based on Graphical Models with Logic for Chinese Named Entity Recognition</i>	
Xiaofeng Yu, Wai Lam and Shing-Kit Chan.....	335
<i>A Hybrid Feature Set based Maximum Entropy Hindi Named Entity Recognition</i>	
Sujan Kumar Saha, Sudeshna Sarkar and Pabitra Mitra.....	343

Relation Extraction

<i>An Effective Methods of using Web based Information for Relation Extraction</i>	
Stanley, Wai Keong Yong and Jian Su.....	350
<i>Minimally Supervised Learning of Semantic Knowledge from Query Logs</i>	
Mamoru Komachi and Hisami Suzuki.....	358

Statistical Models/Machine Learning for NLP (1)

<i>Learning a Stopping Criterion for Active Learning for Word Sense Disambiguation and Text Classification</i>	
Jingbo Zhu, Huizhen Wang and Eduard Hovy.....	366
<i>Multi-View Co-Training of Transliteration Model</i>	
Jin-Shea Kuo and Haizhou Li.....	373
<i>Identifying Sections in Scientific Abstracts using Conditional Random Fields</i>	
Kenji Hirohata, Naoaki Okazaki, Sophia Ananiadou and Mitsuru Ishizuka.....	381

Ontologies and Linguistic Resources (1)

<i>Formalising Multi-layer Corpora in OWL DL - Lexicon Modelling, Querying and Consistency Control</i>	
Aljoscha Burchardt, Sebastian Padó, Dennis Spohr, Anette Frank and Ulrich Heid.....	389
<i>Constructing Taxonomy of Numerative Classifiers for Asian Languages</i>	
Kiyoaki Shirai, Takenobu Tokunaga, Chu-Ren Huang, Shu-Kai Hsieh, Tzu-Yi Kuo, Virach Sornlertlamvanich and Thatsanee Charoenporn.....	397
<i>Translating Compounds by Learning Component Gloss Translation Models via Multiple Languages</i>	
Nikesh Garera and David Yarowsky.....	403

Question Answering

<i>Answering Definition Questions via Temporally-Anchored Text Snippets</i>	
Marius Pasca.....	411
<i>Corpus-based Question Answering for why-Questions</i>	
Ryuichiro Higashinaka and Hideki Isozaki.....	418
<i>Cluster-Based Query Expansion for Statistical Question Answering</i>	
Lucian Vlad Lita and Jaime Carbonell	426

Statistical Models/Machine Learning for NLP (2)

<i>A Semantic Feature for Relation Recognition Using a Web-based Corpus</i>	
Chen-Ming Hung.....	434

<i>Multilingual Text Entry using Automatic Language Detection</i>	
Yo Ehara and Kumiko Tanaka-Ishii.....	441
<i>Using Contextual Speller Techniques and Language Modeling for ESL Error Correction</i>	
Michael Gamon, Jianfeng Gao, Chris Brockett, Alexandre Klementiev, William B. Dolan, Dmitriy Belenko and Lucy Vanderwende.....	449
Ontologies and Linguistic Resources (2)	
<i>Bilingual Synonym Identification with Spelling Variations</i>	
Takashi Tsunakawa and Jun'ichi Tsujii.....	457
<i>Minimally Supervised Multilingual Taxonomy and Translation Lexicon Induction</i>	
Nikesh Garera and David Yarowsky.....	465
<i>Japanese-Spanish Thesaurus Construction Using English as a Pivot</i>	
Jessica Ramirez, Masayuki Asahara and Yuji Matsumoto.....	473
Event/Sentence Relation	
<i>Automatic Identification of Rhetorical Roles using Conditional Random Fields for Legal Document Summarization</i>	
M. Saravanan, B. Ravindran and S. Raman.....	481
<i>Projection-based Acquisition of a Temporal Labeller</i>	
Kathrin Spreyer and Anette Frank.....	489
<i>Acquiring Event Relation Knowledge by Learning Cooccurrence Patterns and Fertilizing Cooccurrence Samples with Verbal Nouns</i>	
Shuya Abe, Kentaro Inui and Yuji Matsumoto.....	497
Statistical Machine Translation	
<i>Refinements in BTG-based Statistical Machine Translation</i>	
Deyi Xiong, Min Zhang, AiTi Aw, Haitao Mi, Qun Liu and Shouxun Lin.....	505
<i>Simple Syntactic and Morphological Processing Can Help English-Hindi Statistical MT</i>	
Ananthakrishnan Ramanathan, Jayprasad Hegde, Ritesh M. Shah, Pushpak Bhattacharyya and Sasikumar M.....	513
<i>Statistical Translation Models for Personalized Search</i>	
Rohini U, Vamshi Ambati and Vasudeva Varma.....	521

Ontologies and Linguistic Resources (3)

<i>Repurposing Theoretical Linguistic Data for Tool Development and Search</i> Fei Xia and William D. Lewis.....	529
<i>Computing Paraphrasability of Syntactic Variants using Web Snippets</i> Atsushi Fujita and Satoshi Sato.....	537
<i>Augmenting Wikipedia with Named Entity Tags</i> Wisam Dakka and Silviu Cucerzan.....	545

Semantic Similarity

<i>Context Feature Selection for Distributional Similarity</i> Masato Hagiwara, Yasuhiro Ogawa and Katsuhiko Toyama.....	553
<i>Gloss-Based Semantic Similarity Metrics for Predominant Sense Acquisition</i> Ryu Iida, Diana McCarthy and Rob Koeling.....	561
<i>Benchmarking Noun Compound Interpretation</i> Su Nam Kim and Timothy Baldwin.....	569

Volume II

Poster papers (Poster session 1)

Vaakkriti: Sanskrit Tokenizer

Aasish Pappu and Ratna Sanyal.....577

A Bottom Up approach to Persian Stemming

Amir Azim Sharifloo and Mehrnoush Shamsfard.....583

Named Entity Recognition in Bengali: A Conditional Random Field Approach

Asif Ekbal, Rejwanul Haque and Sivaji Bandyopadhyay.....589

An Online Cascaded Approach to Biomedical Named Entity Recognition

Shing-Kit Chan, Wai Lam and Xiaofeng Yu.....595

Automatic rule acquisition for Chinese intra-chunk relations

Qiang Zhou.....601

Japanese Named Entity Recognition Using Structural Natural Language Processing

Ryohei Sasano and Sadao Kurohashi.....607

Dimensionality Reduction with Multilingual Resource

YingJu Xia, Hao Yu and Gang Zou.....613

A Web-based English Proofing System for English as a Second Language Users

Xing YI, Jianfeng Gao and William B. Dolan.....619

Analysis of Intention in Dialogues Using Category Trees and Its Application to Advertisement Recommendation

Hung-Chi Huang, Hsin-Hsi Chen and Ming-Shun Lin.....625

Term Extraction Through Unithood and Termhood Unification

Thuy Vu, AiTi Aw and Min Zhang.....631

Search Result Clustering Using Label Language Model

Yeha Lee, Seung-Hoon Na and Jong-Hyeok Lee.....637

Effects of Related Term Extraction in Transliteration into Chinese

HaiXiang Huang and Atsushi Fujii.....643

<i>A Structured Prediction Approach for Statistical Machine Translation</i>	
Dakun Zhang, Le Sun and Wenbo Li.....	649
<i>Method of Selecting Training Data to Build a Compact and Efficient Translation Model</i>	
Keiji Yasuda, Ruiqiang Zhang, Hirofumi Yamamoto and Eiichiro Sumita.....	655
<i>Large and Diverse Language Models for Statistical Machine Translation</i>	
Holger Schwenk and Philipp Koehn.....	661
<i>A linguistic and navigational knowledge approach to text navigation</i>	
Javier Couto and Jean-Luc Minel.....	667
<i>Synset Assignment for Bi-lingual Dictionary with Limited Resource</i>	
Virach Sornlertlamvanich, Thatsanee Charoenporn, Chumpol Mokarat, Hitoshi Isahara, Hammam Riza and Purev Jaimai.....	673
<i>Ranking words for building a Japanese defining vocabulary</i>	
Tomoya Noro and Takehiro Tokuda.....	679
<i>Automatically Identifying Computationally Relevant Typological Features</i>	
William D. Lewis and Fei Xia.....	685
<i>Automatic Paraphrasing of Japanese Functional Expressions Using a Hierarchically Organized Dictionary</i>	
Suguru Matsuyoshi and Satoshi Sato.....	691
<i>Generation of Referring Expression Using Prefix Tree Structure</i>	
Sibabrata Paladhi and Sivaji Bandyopadhyay.....	697
<i>Coverage-based Evaluation of Parser Generalizability</i>	
Tuomo Kakkonen and Erkki Sutinen.....	703
<i>Learning Reliability of Parses for Domain Adaptation of Dependency Parsing</i>	
Daisuke Kawahara and Kiyotaka Uchimoto.....	709
<i>Resolving Ambiguities of Chinese Conjunctive Structures by Divide-and-conquer Approaches</i>	
Duen-Chi Yang, Yu-Ming Hsieh and Keh-Jiann Chen.....	715
<i>Dependency Annotation Scheme for Indian Languages</i>	
Rafiya Begum, Samar Husain, Arun Dhawaj, Dipti Misra Sharma, Lakshmi Bai and Rajeev Sangal.....	721

<i>Non-Factoid Japanese Question Answering through Passage Retrieval that Is Weighted Based on Types of Answers</i>	
Masaki Murata, Sachiyo Tsukawaki, Toshiyuki Kanamaru, Qing Ma and Hitoshi Isahara.....	727
<i>A Multi-Document Multi-Lingual Automatic Summarization System</i>	
Mohamad Ali Honarpisheh, Gholamreza Ghassem-Sani and Ghassem Mirroshandel.....	733
<i>Summarization by Analogy: An Example-based Approach for News Articles</i>	
Megumi Makino and Kazuhide Yamamoto.....	739
<i>Sentence Ordering based on Cluster Adjacency in Multi-Document Summarization</i>	
Donghong Ji and Yu Nie.....	745
<i>Statistical Machine Translation based Passage Retrieval for Cross-Lingual Question Answering</i>	
Tomoyosi Akiba, Kei Shimizu and Atsushi Fujii.....	751
Poster papers (Poster session 2)	
<i>Unsupervised All-words Word Sense Disambiguation with Grammatical Dependencies</i>	
Vivi Nastase.....	757
<i>Syntactic and Semantic Frames in PrepNet</i>	
Saint-Dizier Patrick.....	763
<i>Automatic Classification of English Verbs Using Rich Syntactic Features</i>	
Lin Sun, Anna Korhonen and Yuval Krymolowski.....	769
<i>MRD-based Word Sense Disambiguation: Further Extending Lesk</i>	
Timothy Baldwin, Su Nam Kim, Francis Bond, Sanae Fujita, David Martinez and Takaaki Tanaka.....	775
<i>Fast Computing Grammar-driven Convolution Tree Kernel for Semantic Role Labeling</i>	
Wanxiang Che, Min Zhang, AiTi Aw, Chew Lim Tan, Ting Liu and Sheng Li.....	781
<i>SYNGRAPH: A Flexible Matching Method based on Synonymous Expression Extraction from an Ordinary Dictionary and a Web Corpus</i>	
Tomohide Shibata, Michitaka Odani, Jun Harashima, Takashi Oonishi and Sadao Kurohashi.....	787

<i>Annotation of Multiword Expressions in the Prague Dependency Treebank</i>	
Eduard Bejček, Pavel Straňák and Pavel Schlesinger.....	793
<i>Learning Named Entity Hyponyms for Question Answering</i>	
Paul McNamee, Rion Snow, Patrick Schone and James Mayfield.....	799
<i>Errgrams -- A Way to Improving ASR for Highly Inflected Dravidian Languages</i>	
Kamadev Bhanuprasad and Mats Svenson.....	805
<i>Noise as a Tool for Spoken Language Identification</i>	
Sunita Maithani and J.S. Rawat.....	811
<i>Identifying Real or Fake Articles: Towards better Language Modeling</i>	
Sameer Badaskar, Sachin Agarwal and Shilpa Arora.....	817
<i>Multi-label Text Categorization with Model Combination based on F1-score Maximization</i>	
Akinori Fujino, Hideki Isozaki and Jun Suzuki.....	823
<i>An Experimental Comparison of the Voted Perceptron and Support Vector Machines in Japanese Analysis Tasks</i>	
Manabu Sassano.....	829
<i>Learning Decision Lists with Known Rules for Text Mining</i>	
Venkatesan Chakravarthy, Sachindra Joshi, Ganesh Ramakrishnan, Shantanu Godbole and Sreeram Balakrishnan.....	835
<i>A re-examination of dependency path kernels for relation extraction</i>	
Mengqiu Wang.....	841
<i>Mining Chinese-English Parallel Corpora from the Web</i>	
Bo Li and Juan Liu.....	847
<i>Fast Duplicated Documents Detection using Multi-level Prefix-filter</i>	
Kenji Tateishi and Dai Kusui.....	853
<i>Towards Data And Goal Oriented Analysis: Tool Inter-Operability And Combinatorial Comparison</i>	
Yoshinobu Kano, Ngan Nguyen, Rune Sætre, Kazuhiro Yoshida, Keiichiro Fukamachi, Yusuke Miyao, Yoshimasa Tsuruoka, Sophia Ananiadou and Jun'ichi Tsujii.....	859

<i>A Co-occurrence Graph-based Approach for Personal Name Alias Extraction from Anchor Texts</i> Danushka Bollegala, Yutaka Matsuo and Mitsuru Ishizuka.....	865
<i>Towards Automated Semantic Analysis on Biomedical Research Articles</i> Donghui Feng, Gully Burns, Jingbo Zhu and Eduard Hovy.....	871
<i>Large Scale Diagnostic Code Classification for Medical Patient Records</i> Lucian Vlad Lita, Shipeng Yu, Stefan Niculescu and Jinbo Bi.....	877
<i>Hacking Wikipedia for Hyponymy Relation Acquisition</i> Asuka Sumida and Kentaro Torisawa.....	883
<i>A Discriminative Approach to Japanese Abbreviation Extraction</i> Naoaki Okazaki, Mitsuru Ishizuka and Jun'ichi Tsujii.....	889
<i>Linguistic Interpretation of Emotions for Affect Sensing from Text</i> Mostafa Al Masum Shaikh, Helmut Prendinger and Mitsuru Ishizuka.....	895
<i>How to Take Advantage of the Limitations with Markov Clustering?--The Foundations of Branching Markov Clustering (BMCL)</i> Hiroyuki Akama, Maki Miyake and Jaeyoung Jung.....	901
<i>Combining Context Features by Canonical Belief Network for Chinese Part-Of-Speech Tagging</i> Hongzhi Xu and Chunping Li.....	907
<i>Language Independent Text Correction using Finite State Automata</i> Ahmed Hassan, Sara Noeman and Hany Hassan.....	913
<i>Semantic Role Labeling of Chinese Using Transductive SVM and Semantic Heuristics</i> Yaodong Chen, Ting Wang, Huowang Chen and Xishan Xu.....	919
<i>A Comparative Study of Mixture Models for Automatic Topic Segmentation of Multiparty Dialogues</i> Maria Georgescu, Alexander Clark and Susan Armstrong.....	925
<i>Exploiting Unlabeled Text to Extract New Words of Different Semantic Transparency for Chinese Word Segmentation</i> Richard Tzong-Han Tsai and Hsi-Chuan Hung.....	931

Tutorial

<i>Social Network Inspired Models of NLP and Language Evolution</i> Monojit Choudhury, Animesh Mukherjee and Niloy Ganguly.....	937
<i>How to Add a New Language on the NLP Map: Building Resources and Tools for Languages with Scarce Resources</i> Rada Mihalcea and Vivi Nastase.....	938
<i>Introduction to Text Summarization and Other Information Access Technologies</i> Horacio Saggion.....	939

Demo

<i>A Punjabi Grammar Checker</i> Mandeep Singh Gill, Gurpreet Singh Lehal, and Shiv Sharma Joshi.....	940
<i>Netgraph – Making Searching in Treebanks Easy</i> Jiří Mirovský.....	945
<i>Global Health Monitor - A Web-based System for Detecting and Mapping Infectious Diseases</i> Son Doan, QuocHung-Ngo, Ai Kawazoe, Nigel Collier.....	951
<i>A Mechanism to Provide Language-Encoding Support and an NLP Friendly Editor</i> Anil Kumar Singh.....	957
<i>NLP Applications of Sinhala: TTS & OCR</i> Ruvan Weerasinghe, Asanka Wasala, Dulip Herath and Viraj Welgama.....	963
<i>POLLY: A Conversational System that uses a Shared Representation to Generate Action and Social Language</i> Swati Gupta, Marilyn A. Walker, and Daniela M. Romano.....	967
<i>Cross Lingual Information Access System for Indian Languages</i> CLIA Consortium.....	973
AUTHOR INDEX.....	976

