From the Conference Chair

The Second International Joint Conference on Natural Language Processing (IJCNLP 2005) was prepared to serve as a linking device for advanced research cooperation as well as a sharing device under the AFNLP (Asian Federation of NLP) flag and among worldwide NLP communities.

The twenty-first century, introducing a new international economic world, is experiencing a leap forward in the world of knowledge. The technology of NLP has been achieving rapid growth through the accumulated industrialization experiences and vivid deeper research. We believe this progress is not only contributing to the knowledge society, but also playing an important role in economic growth through awakening the infrastructure of knowledge of the human brain and language functions.

Following the success of IJCNLP 2004, IJCNLP 2005 has made further progress. Not only do we have a big submission increase, but also more invited talks and more tutorials are organized. All these could not have happened without a lot of effort from the conference committees. I’d like to take this opportunity to express my sincere gratitude to the organizing chair Jong-Hyeok Lee, the program co-chairs Robert Dale and Kam-Fai Wong, the publication co-chairs Jian Su and Oi Yee Kwong, and all the other committee chairs for supporting IJCNLP 2005 enthusiastically. It is also my pleasure to thank the AFNLP President Benjamin T’sou, the Vice-President Jun’ichi Tsujii, and the Conference Coordination Committee chair Keh-Yih Su for their continuous advice.

Furthermore, I greatly appreciate the support from the sponsors of this conference: Jeju Province Local Government, KAIST, KISTI, ETRI, Microsoft Korea, Microsoft Japan, and Mobico & Sysmeta.

Last, we look forward to the active participation from you, the honorable guests, to make this conference a successful event.

October 2005
Key-Sun Choi
Conference Chair
IJCNLP 2005
Preface

The Theme of IJCNLP 2005:
“NLP with Kimchee”, a Conference with a Unique Flavor

Welcome to IJCNLP 2005, the second annual conference of the Asian Federation of Natural Language Processing (AFNLP). Following the success of the first conference held in the beautiful city of Sanya, Hainan Island, China, in March 2004, IJCNLP 2005 is held in yet another attractive Asian resort, namely Jeju Island in Korea, on October 11–13, 2005 — the ideal place and season for appreciating mugunghwa, the rose of Sharon, and the national flower of Korea.

On behalf of the Program Committee, we are excited to present these proceedings, which collect together the papers accepted for oral presentation at the conference. We received 289 submissions in total, from 32 economies all over the world: 77% from Asia, 11% from Europe, 0.3% from Africa, 1.7% from Australasia and 10% from North America. We are delighted to report that the popularity of IJCNLP has significantly increased this year, with an increase of 37% from the 211 submissions from 16 economies and 3 continents received for IJCNLP 2004.

With such a large number of submissions, the paper selection process was not easy. With the very considerable assistance of our 12 area chairs — Claire Gardent, Jamie Henderson, Chu-Ren Huang, Kentaro Inui, Gary Lee, Kim-Teng Lua, Helen Meng, Diego Mollá, Jian-Yun Nie, Dragomir Radev, Manfred Stede, and Ming Zhou — and the 133 international reviewers, 90 papers (31%) were accepted for oral presentation and 62 papers (21%) were recommended as posters. The accepted papers were then assigned to 27 parallel sessions leading to a very solid three-day technical program. Four invited speeches were added to further strengthen the program; we are honored this year to have invited Bill Dolan (USA), Seyoung Park (Korea), Karen Spärck Jones (UK) and Hozumi Tanaka (Japan), all world-renowned researchers in their areas, to present their views on the state of the art in natural language processing and information retrieval.

IJCNLP 2005 is obviously a sizable international conference. The work of the Program Committee would not have been so smooth without the unfailing support of the other team members. In particular, we would like to thank Key-Sun Choi (General Chair) and Jong-Hyeok Lee (Organizing Committee Chair) for their continuous backing and their prompt responses to our numerous requests for information and assistance. Thanks are due to the Publication Committee Co-chairs, Jian Su and Oi Yee Kwong, for serving as the bridge between Springer, the authors, and the Program Committee. We are also very grateful to Benjamin T’sou, the President of AFNLP, for his invaluable advice at various stages in the development of the conference.
We hope you will take advantage of every aspect of IJCNLP 2005: the program and the presentations; the proceedings and the papers; the meetings and the people; the resort and the mugunghwa, as well as the food — especially the kimchee. Enjoy it 😊.
Conference Organization

Conference Chair  Key-Sun Choi
                   (KAIST, Korea)
Organizing Chair  Jong-Hyeok Lee
                   (POSTECH, Korea)
Program Co-chairs  Robert Dale
                   (Macquarie University, Australia)
                   Kam-Fai Wong
                   (Chinese University of Hong Kong, Hong Kong)
Publication Co-chairs  Jian Su
                    (Institute for Infocomm Research, Singapore)
                    Oi Yee Kwong
                    (City University of Hong Kong, Hong Kong)
Publicity Co-chairs  Hiroshi Nakagawa
                    (University of Tokyo, Japan)
                    Jong C. Park
                    (KAIST, Korea)
Financial Co-chairs  Hyeok-Cheol Kwon
                    (Busan National University, Korea)
                    Takenobu Tokunaga
                    (Tokyo Institute of Technology, Japan)
Poster and Demo Co-chairs  Rajeev Sangal
                           (IIIT, India)
                           Dekang Lin
                           (Google, USA)
                           Maosong Sun
                           (Tsinghua University, China)
Tutorial Chair  Dekai Wu
                  (HKUST, Hong Kong)
Workshop Co-chairs  Yuji Matsumoto
                   (NAIST, Japan)
                   Laurent Romary
                   (LORIA, France)
Exhibition Co-chairs  Seung-Shik Kang
                     (Kookmin University, Korea)
                     Tetsuya Ishikawa
                     (University of Tsukuba, Japan)
Program Committee

Program Co-chairs

Robert Dale, Macquarie University, Australia
Kam-Fai Wong, Chinese University of Hong Kong, Hong Kong

Area Chairs

Claire Gardent, CNRS/LORIA, Nancy, France
  (Dialogue and Discourse)
James Henderson, University of Geneva, Switzerland
  (Parsing and Grammatical Formalisms)
Chu-Ren Huang, Academia Sinica, Taiwan
  (Semantics and Ontology)
Kentaro Inui, Nara Institute of Science and Technology, Japan
  (Text and Sentence Generation)
Gary Geumbae Lee, POSTECH, Korea
  (Text Mining and Information Extraction)
Kim-Teng Lua, COLIPS, Singapore
  (POS Tagging, WSD and Word Segmentation)
Helen Meng, Chinese University of Hong Kong, Hong Kong
  (Spoken Language Processing)
Diego Mollá, Macquarie University, Australia
  (Question Answering)
Jian-Yun Nie, University of Montreal, Canada
  (Information Retrieval)
Dragomir Radev, University of Michigan, USA
  (Text Summarization and Opinion Extraction)
Manfred Stede, University of Potsdam, Germany
  (Machine Translation)
Ming Zhou, Microsoft Research Asia, China
  (NLP Applications)

Reviewers

Jan Alexandersson  Paul Buitelaar  John Chen
John Bateman  Harry Bunt  Ruey-Cheng Chen
Farah Benamara  John Carroll  Pu-Jen Cheng
Emily Bender  Yee Seng Chan  Lee-Feng Chien
Kalina Bontcheva  Berlin Chen  Key-Sun Choi
Johan Bos  Hsin-Hsi Chen  Charles Clarke
Eric de la Clergerie
Mike Collins
Robert Dale
Angelo Dalli
Thierry Declerck
Minghui Dong
Mark Dras
Terumasa Ehara
Christiane Fellbaum
Guo Hong Fu
Jianfeng Gao
Claire Gardent
Ulrich Germann
Sanda Harabagiu
Donna Harman
James Henderson
Graeme Hirst
Julia Hockenmaier
Chu-Ren Huang
Kentaro Inui
Shuichi Itahashi
Dong Hong Ji
Hongyan Jing
Noriko Kando
Hideki Kashioka
Tatsuya Kawahara
Frank Keller
Jin-dong Kim
Alistair Knott
Wessel Kraaij
Sadao Kurohashi
Oi Yee Kwong
Wai Lam
Helen Langone
Mirella Lapata
Gary Geunbae Lee
Alessandro Lenci
Gina Levow
Haizhou Li
Mu Li
Wei Li
Wenjie Li
Elizabeth Liddy
Chin-Yew Lin
Dekang Lin
Charles Ling
Qun Liu
Ting Liu
Yang Liu
Wai-Kit Lo
Vincenzo Lombardo
Jin Kiat Low
Robert Luk
Yajuan Lv
Qing Ma
Bernardo Magnini
Yuval Marom
Yuji Matsumoto
Kathleen McCoy
Helen Meng
Paola Merlo
Diego Mollá
Josiane Mothe
Dragos Munteanu
Mark-Jan Nederhof
Ani Nenkova
Hwee Tou Ng
Jian-Yun Nie
Takashi Ninomiya
Douglas Oard
Kiyonori Ohtake
Alessandro Oltramari
Miles Osborne
Patrick Paroubek
Adam Pease
Gerald Penn
Massimo Poesio
Dragomir Radev
Deepak Ravichandran
Fuji Ren
Fabio Rinaldi
Rajeev Sangal
Rolf Schwitter
James Shaw
Claude Shen
Melanie Siegel
Khalil Sima’an
Virach Sorlertlamvanich
Manfred Stede
Kristina Striegnitz
Maosong Sun
Richard Sutcliffe
Bing Swen
Simone Teufel
Thanaruk Theeramunkong
Takenobu Tokunaga
Jun’ichi Tsujii
Yoshimasa Tsuruoka
Gertjan van Noord
Jose Luis Vicedo
Chao Wang
Hsin-Min Wang
Hui Wang
Jenq-Haur Wang
Bonnie Webber
David Weir
Fuliang Weng
Janyce Wiebe
Florian Wolf
Kam-Fai Wong
Yunqing Xia
Jie Xu
Kazuhide Yamamoto
Chris Yang
Hongyuan Zha
ChengXiang Zhai
Weidong Zhan
Jie Zhang
Min Zhang
Yujie Zhang
Zhu Zhang
Tiejun Zhao
GuoDong Zhou
Liang Zhou
Ming Zhou
Ingrid Zukerman
Sponsors

Jeju Province Local Government
Korea Advanced Institute of Science and Technology (KAIST)
Korea Institute of Science and Technology Information (KISTI)
Electronics and Telecommunications Research Institute (ETRI)
Microsoft Korea
Microsoft Japan
Mobico & Sysmeta
Table of Contents

Information Retrieval

A New Method for Sentiment Classification in Text Retrieval
   Yi Hu, Jianyong Duan, Xiaoming Chen, Bingzhen Pei, Ruzhan Lu .................................................. 1

Topic Tracking Based on Linguistic Features
   Fumiyo Fukumoto, Yusuke Yamaji ........................................ 10

The Use of Monolingual Context Vectors for Missing Translations in Cross-Language Information Retrieval
   Yan Qu, Gregory Grefenstette, David A. Evans ......................... 22

Automatic Image Annotation Using Maximum Entropy Model
   Wei Li, Maosong Sun ...................................................... 34

Corpus-Based Parsing

Corpus-Based Analysis of Japanese Relative Clause Constructions
   Takeshi Abekawa, Manabu Okumura ..................................... 46

Parsing Biomedical Literature
   Matthew Lease, Eugene Charniak ......................................... 58

Parsing the Penn Chinese Treebank with Semantic Knowledge
   Deyi Xiong, Shuanglong Li, Qin Liu, Shouxun Lin, Yueliang Qian ....................................................... 70

Using a Partially Annotated Corpus to Build a Dependency Parser for Japanese
   Manabu Sassano ............................................................. 82

Web Mining

Entropy as an Indicator of Context Boundaries: An Experiment Using a Web Search Engine
   Kumiko Tanaka-Ishii ...................................................... 93

Automatic Discovery of Attribute Words from Web Documents
   Kosuke Tokunaga, Jun’ichi Kazama, Kentaro Torisawa ............... 106
Aligning Needles in a Haystack: Paraphrase Acquisition Across the Web
Marius Pašca, Péter Dienes ........................................ 119

Confirmed Knowledge Acquisition Using Mails Posted to a Mailing List
Yasuhiko Watanabe, Ryo Nishimura, Yoshihiro Okada .............. 131

Rule-Based Parsing

Automatic Partial Parsing Rule Acquisition Using Decision Tree Induction
Myung-Seok Choi, Chul Su Lim, Key-Sun Choi .................... 143

Chunking Using Conditional Random Fields in Korean Texts
Yong-Hun Lee, Mi-Young Kim, Jong-Hyeok Lee .................... 155

High Efficiency Realization for a Wide-Coverage Unification Grammar
John Carroll, Stephan Oepen ........................................ 165

Linguistically-Motivated Grammar Extraction, Generalization and Adaptation
Yu-Ming Hsieh, Duen-Chi Yang, Keh-Jiann Chen ................... 177

Disambiguation

PP-Attachment Disambiguation Boosted by a Gigantic Volume of Unambiguous Examples
Daisuke Kawahara, Sadao Kurohashi ............................. 188

Adapting a Probabilistic Disambiguation Model of an HPSG Parser to a New Domain
Tadayoshi Hara, Yusuke Miyao, Jun’ichi Tsujii .................... 199

A Hybrid Approach to Single and Multiple PP Attachment Using WordNet
Akshar Bharathi, Rohini U., Vishnu P., S.M. Bendre,
Rajeev Sangal ......................................................... 211

Period Disambiguation with Maxent Model
Chunyu Kit, Xiaoyue Liu ............................................ 223

Text Mining

Acquiring Synonyms from Monolingual Comparable Texts
Mitsuo Shimohata, Eiichiro Sumita ............................... 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

*Gumwon Hong* .......................................................... 366

Discovering Relations Between Named Entities from a Large Raw Corpus Using Tree Similarity-Based Clustering

*Min Zhang, Jian Su, Danmei Wang, Guodong Zhou, Chew Lim Tan* ................................. 378

Automatic Relation Extraction with Model Order Selection and Discriminative Label Identification

*Chen Jinxiu, Ji Donghong, Tan Chew Lim, Niu Zhengyu* ............ 390

Mining Inter-Entity Semantic Relations Using Improved Transductive Learning

*Zhu Zhang* .............................................................. 402

## Text Classification

A Preliminary Work on Classifying Time Granularities of Temporal Questions

*Wei Li, Wenjie Li, Qin Lu, Kam-Fai Wong* ........................................ 414

Classification of Multiple-Sentence Questions

*Akihiro Tamura, Hiroya Takamura, Manabu Okumura* ..................... 426

## Transliteration

A Rule Based Syllabification Algorithm for Sinhala

*Ruvan Weerasinghe, Asanka Wasala, Kumudu Gamage* .................... 438

An Ensemble of Grapheme and Phoneme for Machine Transliteration

*Jong-Hoon Oh, Key-Sun Choi* ............................................. 450

## Machine Translation – I

Improving Statistical Word Alignment with Ensemble Methods

*Wu Hua, Wang Haifeng* .................................................. 462

Empirical Study of Utilizing Morph-Syntactic Information in SMT

*Young-Sook Hwang, Taro Watanabe, Yutaka Sasaki* ...................... 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, Yüji 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
Min Zhang, Haizhou Li, Jian Su, Hendra Setiawan ................. 600

Machine Translation Based on Constraint-Based Synchronous Grammar
Fai Wong, Dong-Cheng Hu, Yu-Hang Mao, Ming-Chui Dong, Yi-Ping Li........................................... 612

Text Summarization

A Machine Learning Approach to Sentence Ordering for Multidocument Summarization and Its Evaluation
Danushka Bollegala, Naoaki Okazaki, Mitsuru Ishizuka ........... 624

Significant Sentence Extraction by Euclidean Distance Based on Singular Value Decomposition
Changbeom Lee, Hyukro Park, Cheolyoung Ock ..................... 636

Named Entity Recognition

Two-Phase Biomedical Named Entity Recognition Using a Hybrid Method
Seonho Kim, Juntae Yoon, Kyung-Mi Park, Hae-Chang Rim ...... 646

Heuristic Methods for Reducing Errors of Geographic Named Entities Learned by Bootstrapping
Seungwoo Lee, Gary Geunbae Lee ..................................... 658

Linguistic Resources and Tools

Building a Japanese-Chinese Dictionary Using Kanji/Hanzi Conversion
Chooi-Ling Goh, Masayuki Asahara, Yuji Matsumoto ............... 670

Automatic Acquisition of Basic Katakana Lexicon from a Given Corpus
Toshiaki Nakazawa, Daisuke Kawahara, Sadao Kurohashi .......... 682

CTEMP: A Chinese Temporal Parser for Extracting and Normalizing Temporal Information
Wu Mingli, Li Wenjie, Lu Qin, Li Baoli ................................. 694

French-English Terminology Extraction from Comparable Corpora
Béatrice Daille, Emmanuel Morin .................................... 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 Isu .................................................... 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  
Conrad Chen, Hsin-Hsi Chen ........................................... 838

A Connectionist Model of Anticipation in Visual Worlds  
Marshall R. Mayberry, III, Matthew W. Crocker, Pia Knoeferle ...... 849

Tagging

Automatically Inducing a Part-of-Speech Tagger by Projecting from Multiple Source Languages Across Aligned Corpora  
Victoria Fossum, Steven Abney .......................................... 862

The Verbal Entries and Their Description in a Grammatical Information-Dictionary of Contemporary Tibetan  
Jiang Di, Long Congjun, Zhang Jichuan ................................. 874

Tense Tagging for Verbs in Cross-Lingual Context: A Case Study  
Yang Ye, Zhu Zhang .......................................................... 885

Regularisation Techniques for Conditional Random Fields: Parameterised Versus Parameter-Free  
Andrew Smith, Miles Osborne ............................................. 896

Semantic Analysis – II

Exploiting Lexical Conceptual Structure for Paraphrase Generation  
Atsushi Fujita, Kentaro Inui, Yuji Matsumoto ............................ 908

Word Sense Disambiguation by Relative Selection  
Hee-Cheol Seo, Hae-Chang Rim, Myung-Gil Jang ....................... 920

Towards Robust High Performance Word Sense Disambiguation of English Verbs Using Rich Linguistic Features  
Jinying Chen, Martha Palmer .............................................. 933

Automatic Interpretation of Noun Compounds Using WordNet Similarity  
Su Nam Kim, Timothy Baldwin ............................................. 945

Language Models

An Empirical Study on Language Model Adaptation Using a Metric of Domain Similarity  
Wei Yuan, Jianfeng Gao, Hisami Suzuki .................................. 957
A Comparative Study of Language Models for Book and Author Recognition  
Özlem Uzuner, Boris Katz .................................................. 969

Spoken Language

Lexical Choice via Topic Adaptation for Paraphrasing Written Language to Spoken Language  
Nobuhiro Kaji, Sadao Kurohashi ................................. 981

A Case-Based Reasoning Approach for Speech Corpus Generation  
Yandong Fan, Elizabeth Kendall ....................................... 993

Terminology Mining

Web-Based Terminology Translation Mining  
Gaolin Fang, Hao Yu, Fumihito Nishino ......................... 1004

Extracting Terminologically Relevant Collocations in the Translation of Chinese Monograph  
Byeong-Kwu Kang, Bao-Bao Chang, Yi-Rong Chen,  
Shi-Wen Yu ........................................................... 1017

Author Index ............................................................ 1029