

# **2009 International Conference on Asian Language Processing**

**(IALP 2009)**

**Singapore  
7-9 December 2009**



**IEEE Catalog Number: CFP0944I-PRT**  
**ISBN: 978-1-4244-5318-4**

# 2009 International Conference on Asian Language Processing

## IALP 2009

# Table of Contents

Message from General Chair  
Message from Program Chair  
Conference Committees  
Program Committee  
Invited Talks

---

### IALP 2009

An Articulation Training System with Intelligent Interface and Multimode Feedbacks to Articulation Disorders .....	3
<i>Yeou-Jiunn Chen, Jiunn-Liang Wu, Hui-Mei Yang, Chung-Hsien Wu, Chih-Chang Chen, and Shan-Shan Ju</i>	
Research on Feature Extraction from Chinese Text for Opinion Mining .....	7
<i>Shanzong Zhu, Yuanchao Liu, Ming Liu, and Peiliang Tian</i>	
Dialog-Act Recognition Using Discourse and Sentence Structure Information .....	11
<i>Keyan Zhou and Chengqing Zong</i>	
Multilingual Multimodal Integration of Sketch and Speech: A Generic Speech Representation Model for Spatial Description .....	17
<i>Lee-Na Teh and Alvin W. Yeo</i>	
An Experimental Study on Vietnamese POS Tagging .....	23
<i>Oanh Thi Tran, Cuong Anh Le, Thuy Quang Ha, and Quynh Hoang Le</i>	
Similarity Computing Between Simple Questions Based on HowNet and Syntactic Analysis .....	28
<i>Zhou Zhibin, Shi Shuicai, Han Yanhua, and Lv Xueqiang</i>	
Information Focus Synthesis Based on Question Answer Chain .....	34
<i>Jing Wan and Han Ren</i>	
Stability vs. Effectiveness: Improved Sentence-Level Combination of Machine Translation Based on Weighted MBR .....	39
<i>Bo Wang, Tiejun Zhao, Muyun Yang, Hongfei Jiang, and Sheng Li</i>	

Repetition in Mandarin Interaction: A Case Study on TV Shopping Channels in Taiwan .....	43
<i>Fuhui Hsieh and Ying Liang</i>	
Building Thai WordNet with a Bi-directional Translation Method .....	48
<i>Dhanon Leenoi, Thepchai Supnithi, and Wirote Aroonmanakun</i>	
Segmenting Long Sentence Pairs for Statistical Machine Translation .....	53
<i>Biping Meng, Shujian Huang, Xinyu Dai, and Jiajun Chen</i>	
Combining ILP and MLN for Coreference Resolution .....	59
<i>Yabing Zhang, Junsheng Zhou, Shujian Huang, and Jiajun Chen</i>	
Improved Reordering Rules for Hierarchical Phrase-Based Translation .....	65
<i>Shu Cai, Yajuan Lü, and Qun Liu</i>	
A Hybrid Method of Chinese Prosodic Word Tagging Based on Keyword Anchor and Hidden Markov Model .....	71
<i>Zhou Quan, Deng Pan, Liu Hongjian, Guo Defeng, and Nagamatsu Kenji</i>	
An Acoustic Comparison of Vowel Length Contrasts in Standard Arabic, Japanese and Thai .....	76
<i>Kimiko Tsukada</i>	
Semantic Role Based Tamil Sentence Generator .....	80
<i>S. Lakshmana Pandian and T.V. Geetha</i>	
Vietnamese Final Stop Consonants /p, t, k/ Described in Terms of Formant Transition Slopes .....	86
<i>Viet Son Nguyen, Eric Castelli, and René Carré</i>	
Foxinfo1.0: A Chinese Topic-Oriented Search Engine .....	91
<i>Ke Sun, Lei Lin, Bingquan Liu, Chengjie Sun, and Xiaolong Wang</i>	
A Bi-directional Translation Approach for Building Thai Wordnet .....	97
<i>Prissana Akaraputthiporn, Krit Kosawat, and Wirote Aroonmanakun</i>	
A Study on Semantic Role Labeling of Korean Sentence .....	102
<i>Yude Bi and Jie Chen</i>	
Challenges in Developing Persian Corpora from Online Resources .....	108
<i>Masood Ghayoomi and Saeedeh Momtazi</i>	
A Wavelet-Based Denoising System Using Time-Frequency Adaptation for Speech Enhancement .....	114
<i>Kun-Ching Wang, Chuin-Li Chin, and Yi-Hsing Tsai</i>	
Extracting Thai Compounds Using Collocations and POS Bigram Probabilities without a POS Tagger .....	118
<i>Wirote Aroonmanakun</i>	
Analysis and Selection of Prosodic Features for Language Identification .....	123
<i>Raymond W.M. Ng, Tan Lee, Cheung-Chi Leung, Bin Ma, and Haizhou Li</i>	
A New Way for Chinese Place Name Recognition .....	129
<i>Quan Zhang, Yi Yuan, Nuo Li, Xiangfeng Wei, and Jianming Miao</i>	

Improving Chinese to English SMT with Multiple CWS Results .....	135
<i>Yongliang Ma and Tiejun Zhao</i>	
A Modular Cascaded Approach to Complete Parsing .....	141
<i>Samar Husain, Phani Gadde, Bharat Ambati, Dipti Misra Sharma, and Rajeev Sangal</i>	
Automatic User Preferences Acquirement in Chinese Commercial Web Sites with NLP and DM Techniques .....	147
<i>Shilin Zhang and Hui Wang</i>	
An Experimental Study of Vietnamese Question Answering System .....	152
<i>Vu Mai Tran, Vinh Duc Nguyen, Oanh Thi Tran, Uyen Thu Thi Pham, and Thuy Quang Ha</i>	
Privacy Sensitivity: Application in Arabic .....	156
<i>Sabah Al-Fedaghi and Faisal Alhaqan</i>	
Automatic Machine Translation Evaluation Based on Sentence Structure Information .....	162
<i>Ding Ze-ya, Zang Han-fen, Zhang Quan, Miao Jian-ming, and Chi Yu-huan</i>	
Research on the Recognition of fKs Based on HNC Theory .....	167
<i>Hanfen Zang, Xiangfeng Wei, and Quan Zhang</i>	
A Three-Pass System Combination Framework by Combining Multiple Hypothesis Alignment Methods .....	172
<i>Jinhua Du and Andy Way</i>	
The Automatic Analysis of Chunk-Extension Sentences .....	177
<i>Xiangfeng Wei and Quan Zhang</i>	
On the Semantic Orientation and Computer Identification of Adverb bié .....	182
<i>Lin He and Di Wu</i>	
Author Profiling for Vietnamese Blogs .....	190
<i>Dang Duc Pham, Giang Binh Tran, and Son Bao Pham</i>	
Exploring the Effects of Text Clustering on On-Line Military News Based on Quantitative Association Rule .....	195
<i>Liang-chu Chen, Chyi-Bao Yang, Jih-Hsin Chen, and Yen-Hsuan Lien</i>	
Exploring Syntactic Features for Pronoun Resolution Using Context-Sensitive Convolution Tree Kernel .....	201
<i>Fang Kong, Yancui Li, Guodong Zhou, and Qiaoming Zhu</i>	
Research on Text Hierarchical Topic Identification Algorithm Based on the Dynamic Diverse Thresholds Clustering .....	206
<i>Xu Yong-Dong, Quan Guang-Ri, Xu Zhi-Ming, and Wang Ya-Dong</i>	
Automatic Domain-Ontology Relation Extraction from Semi-structured Texts .....	211
<i>Cheng Xiao, Dequan Zheng, Yuhang Yang, and Guojun Shao</i>	
Proofreader of Mongolian Vocabulary Based on Language Model of Syllabic Statistics .....	217
<i>Ochir, Menghjiya, and Gong-Zheng</i>	

Improving the Performance of the Link Parser .....	220
<i>Yalamanchi Viswanatha Naidu, Anil Kumar Singh, Dipti Misra Sharma, and Akshar Bharati</i>	
Chinese Chunking Based on Coarse-Grained Part-of-Speech Features .....	226
<i>Guang-Lu Sun, Yibo Xue, Zhiming Xu, and Fei Lang</i>	
Identification of Closely Related Indigenous Languages: An Orthographic Approach .....	230
<i>Ee-Lee Ng, Alvin W. Yeo, and Bali Ranaivo-Malancon</i>	
Research on Chinese Text Summarization Algorithm Based on Statistics and Rules .....	236
<i>Faguo Zhou, Fan Zhang, and Bingru Yang</i>	
Two-Pass Deterministic Dependency Parsing for Long Chinese Sentences .....	240
<i>Ping Jian and Chengqing Zong</i>	
Refining Unit Boundaries for Mandarin Text-to-Speech Database .....	245
<i>Minghui Dong, Ling Cen, Paul Chan, and Haizhou Li</i>	
Hedge-Based Filtering Method for Semi-structure Documents .....	249
<i>Tong Wang, Chunhui Zhao, and Lei Yang</i>	
Semantic Genes and the Semantic Composition of Adjectives in Modern Chinese .....	254
<i>Dan Hu and Jinglian Gao</i>	
Approaches of Dimensionality Reduction for Telugu Document Classification .....	259
<i>P. Vijayapal Reddy, B. Sasidhar, B. Harinatha Reddy, B. Vishnu Vardhan, L. Pratap Reddy, and A. Govardhan</i>	
Automatic Chinese Essay Scoring Using Connections between Concepts in Paragraphs .....	265
<i>Tao-Hsing Chang and Chia-Hoang Lee</i>	
Study on the Classification of Mixed Text Based on Conceptual Vector Space Model and Bayes .....	269
<i>Yaxiong Li and Dan Hu</i>	
Semi-supervised Learning of Domain-Specific Language Models from General Domain Data .....	273
<i>Shuanhu Bai, Min Zhang, and Haizhou Li</i>	
Advances in Acoustic Modeling for Vietnamese LVCSR .....	280
<i>Tuan Nguyen and Quan Vu</i>	
A Lattice-Based Phonotactic Language Recognition System with CMLLR Adaptation and Its Implementation Issues .....	285
<i>Cheung-Chi Leung, Rong Tong, Bin Ma, and Haizhou Li</i>	
Chinese Named Entity Recognition Using a Morpheme-Based Chunking Tagger .....	289
<i>Guohong Fu</i>	

Learning Method for Extraction of Partial Correspondence from Parallel Corpus .....	293
<i>Ryo Terashima, Hiroshi Echizen-ya, and Kenji Araki</i>	
A Kappa Measurement of Query Consistency and Its Application .....	299
<i>Chao Song, Muyun Yang, Haoliang Qi, and Sheng Li</i>	
Sogou Query Log Analysis: A Case Study for Collaborative Recommendation or Personalized IR .....	304
<i>Zhitao Zhang, Muyun Yang, Sheng Li, Haoliang Qi, and Chao Song</i>	
Ranking vs. Classification: A Case Study in Mining Organization Name Translation from Snippets .....	308
<i>Muyun Yang, Zhenyong Shi, Sheng Li, Tiejun Zhao, and Haoliang Qi</i>	
The Improved Logistic Regression Models for Spam Filtering .....	314
<i>Yong Han, Muyun Yang, Haoliang Qi, Xiaoning He, and Sheng Li</i>	
Automatic Acquisition of Large-Scale Academic Bilingual Parallel Corpus from the Web .....	318
<i>Han Yong, Li Yu, He Xiaoning, Yang Muyun, and Lei Guohua</i>	
BEST Corpus Development and Analysis .....	322
<i>Monthika Boriboon, Kanyanut Kriengket, Patcharika Chootrakool, Sitthaa Phaholphinyo, Sumonmas Purodakananda, Tipraporn Thanakulwarapas, and Krit Kosawat</i>	
Three Algorithms for Word-to-Phrase Machine Translation .....	328
<i>Le Manh Hai and Phan Thi Tuoi</i>	
A Study on Eye Movement of Korean Students Reading Chinese Texts with or without Marks for Word Boundaries .....	332
<i>Yu Peng and Liu Su</i>	

## Author Index