

2010 International Conference on Asian Language Processing

(IALP 2010)

**Harbin, China
28 – 30 December 2010**



IEEE Catalog Number: CFP1044I-PRT
ISBN: 978-1-4244-9063-9

2010 International Conference on Asian Language Processing *IALP 2010*

Table of Contents

Message from General Chairs	xi
Message from Program Chairs	xiii
Conference Committees	xiv
Program Committee	xv
Organizers and Sponsors	xvii
Invited Talks	xix

Lexicon and Morphology

A Survey on Rendering Traditional Mongolian Script	3
<i>Biligsaikhan Batjargal, Fuminori Kimura, and Akira Maeda</i>	
A Combination of Statistical and Rule-Based Approach for Mongolian Lexical Analysis	7
<i>Lili Zhao, Jia Men, Congpin Zhang, Qun Liu, Wenbin Jiang, Jinxing Wu, and Qing Chang</i>	
A Letter Tagging Approach to Uyghur Tokenization	11
<i>Batuer Aisha</i>	
Development of Analysis Rules for Bangla Root and Primary Suffix for Universal Networking Language	15
<i>Md. Nawab Yousuf Ali, Shahid Al Noor, Md. Zakir Hossain, and Jugal Krishna Das</i>	
A Suffix-Based Noun and Verb Classifier for an Inflectional Language	19
<i>Navanath Saharia, Utpal Sharma, and Jugal Kalita</i>	
Behavior of Word ‘kaa’ in Urdu Language	23
<i>Muhammad Kamran Malik, Aasim Ali, and Shahid Siddiq</i>	
Methods to Divide Uygur Morphemes and Treatments for Exceptions	27
<i>Pu Li and Hao Zhao</i>	
Rules for Morphological Analysis of Bangla Verbs for Universal Networking Language	31
<i>Md. Nawab Yousuf Ali, Mohammad Zakir Hossain Sarker, Ghulam Farooque Ahmed, and Jugal Krishna Das</i>	

Discussion on Collation of Tibetan Syllable	35
<i>Heming Huang and Feipeng Da</i>	
A Dictionary Mechanism for Chinese Word Segmentation Based on the Finite Automata	39
<i>Wu Yang, Liyun Ren, and Rong Tang</i>	
Development of Templates for Dictionary Entries of Bangla Roots and Primary Suffixes for Universal Networking Language	43
<i>Md. Zakir Hossain, Shaikh Muhammad Allayear, Md. Nawab Yousuf Ali, and Jugal Krishna Das</i>	
A Study on "Worry" Separable Words & Its Separable Slots	47
<i>Chunling Li and Xiaoxiao Wang</i>	
Syntax and Parsing	
Improving Dependency Parsing Using Punctuation	53
<i>Zhenghua Li, Wanxiang Che, and Ting Liu</i>	
A Tree Probability Generation Using VB-EM for Thai PGLR Parser	57
<i>Kanokorn Trakultaweekoon, Taneth Ruangrajitpakorn, Prachya Boonkwan, and Thepchai Supnithi</i>	
Research on Verb Subcategorization-Based Syntactic Parsing Postprocess for Chinese Language	61
<i>Jinyong Wang and Xiwu Han</i>	
Identification of Maximal-Length Noun Phrases Based on Maximal-Length Preposition Phrases in Chinese	65
<i>Guiping Zhang, Wenjing Lang, Qiaoli Zhou, and Dongfeng Cai</i>	
Urdu Noun Phrase Chunking - Hybrid Approach	69
<i>Shahid Siddiq, Sarmad Hussain, Aasim Ali, Kamran Malik, and Wajid Ali</i>	
The Function of Fixed Word Combination in Chinese Chunk Parsing	73
<i>Liqun Wang and Shoichi Yokoyama</i>	
Problems and Review of Statistical Parsing Language Model	77
<i>Faguo Zhou, Fan Zhang, and Bingru Yang</i>	
A General Comparison on Sentences Analysis and Its Teaching Significance between Traditional and Structural Grammars	81
<i>Jiaying Yu</i>	
Semantics	
Two Cores in Chinese Negation System: A Corpus-Based View	87
<i>Hio Tong Chan and Chunyu Kit</i>	
Finding Semantic Similarity in Vietnamese	91
<i>Dat Tien Nguyen and Son Bao Pham</i>	
Automatic Metaphor Recognition Based on Semantic Relation Patterns	95
<i>Xuri Tang, Weiguang Qu, Xiaohe Chen, and Shiwen Yu</i>	

Event Entailment Extraction Based on EM Iteration	101
<i>Zhen Li, Hanjing Li, Mo Yu, Tiejun Zhao, and Sheng Li</i>	
On the Semantic Orientation and Computer Identification of the Adverb “Jiù”	105
<i>Lin He and Jiaqin Wu</i>	
Semantic Genes and the Formalized Representation of Lexical Meaning	110
<i>Dan Hu</i>	
Acquisition of Hypernymy-Hyponymy Relation between Nouns for WordNet Building	114
<i>Gunawan and Erick Pranata</i>	
Algorithm for Conversion of Bangla Sentence to Universal Networking Language	118
<i>Md. Nawab Yousuf Ali, M. Ameer Ali, Abu Mohammad Nurannabi, and Jugal Krishna Das</i>	
Construction of the Paradigmatic Semantic Network Based on Cognition	122
<i>Xiaofang Ouyang</i>	
The Research of Sentence Testing Based on HNC Analysis System of Sentence Category	126
<i>Zhiying Liu</i>	
Semantic Patterns of Chinese Post-Modified V+N Phrases	130
<i>Likun Qiu and Wenxian Zhang</i>	
Information Extraction	
A Grammar-Based Unsupervised Method of Mining Volitive Words	137
<i>Jianfeng Zhang, Yu Hong, Yuehui Yang, Jianmin Yao, and Qiaoming Zhu</i>	
Using Feature Selection to Speed Up Online SVM Based Spam Filtering	142
<i>Yuewu Shen, Guanglu Sun, Haoliang Qi, and Xiaoning He</i>	
A Semi-supervised Method for Classification of Semantic Relation between Nominals	146
<i>Yuan Chen, Yue Lu, Man Lan, Jian Su, and Zhengyu Niu</i>	
XPath-Wrapper Induction for Data Extraction	150
<i>Nam-Khanh Tran, Kim-Cuong Pham, and Quang-Thuy Ha</i>	
A Block Segmentation Based Approach for Web Information Extraction	154
<i>Chanwei Wang, Chengjie Sun, Lei Lin, and Xiaolong Wang</i>	
Linguistic Features for Named Entity Recognition Using CRFs	158
<i>R. Vijay Sundar Ram, A. Akilandeswari, and Sobha Lalitha Devi</i>	
Research on Domain-Adaptive Transfer Learning Method and Its Applications	162
<i>Geli Fei and Dequan Zheng</i>	

Information Theory Based Feature Valuing for Logistic Regression for Spam Filtering	166
<i>Haoliang Qi, Xiaoning He, Yong Han, Muyun Yang, and Sheng Li</i>	
Automatic Named Entity Set Expansion Using Semantic Rules and Wrappers for Unary Relations	170
<i>Mai-Vu Tran, Tien-Tung Nguyen, Thanh-Son Nguyen, and Hoang-Quynh Le</i>	
Anaphora Resolution of Malay Text: Issues and Proposed Solution Model	174
<i>Noorhuzaimi Karimah Mohd Noor, Shahrul Azman Noah, Mohd Juzaidin Ab. Aziz, and Mohd Pouzi Hamzah</i>	
Combining Multi-features with Conditional Random Fields for Person Recognition	178
<i>Suxiang Zhang</i>	
Comparison between Typical Discriminative Learning Model and Generative Model in Chinese Short Messages Service Spam Filtering	182
<i>Xiaoxia Zheng, Chao Liu, Chengzhe Huang, Yu Zou, and Hongwei Yu</i>	
Chinese Spam Filter Based on Relaxed Online Support Vector Machine	185
<i>Yong Han, Xiaoning He, Muyun Yang, Haoliang Qi, and Chao Song</i>	
Comment Target Extraction Based on Conditional Random Field & Domain Ontology	189
<i>Shengchun Ding and Ting Jiang</i>	
Text Understanding and Summarization	
Topic-Driven Multi-document Summarization	195
<i>Hongling Wang and Guodong Zhou</i>	
Multiple Factors-Based Opinion Retrieval and Coarse-to-Fine Sentiment Classification	199
<i>Shu Zhang, Wenjie Jia, Yingju Xia, Yao Meng, and Hao Yu</i>	
Chinese Sentence-Level Sentiment Classification Based on Sentiment Morphemes	203
<i>Xin Wang and Guohong Fu</i>	
Extracting Phrases in Vietnamese Document for Summary Generation	207
<i>Huong Thanh Le, Rathany Chan Sam, and Phuc Trong Nguyen</i>	
User Interest Analysis with Hidden Topic in News Recommendation System	211
<i>Mai-Vu Tran, Xuan-Tu Tran, and Huy-Long Uong</i>	
Dependency Tree-Based Anaphoricity Determination for Coreference Resolution	215
<i>Fang Kong, Jianmei Zhou, Guodong Zhou, and Qiaoming Zhu</i>	
Text Clustering Based on Domain Ontology and Latent Semantic Analysis	219
<i>Yaxiong Li, Jianqiang Zhang, and Dan Hu</i>	

Social Network Mining Based on Wikipedia	223
<i>Fangfang Yang, Zhiming Xu, Sheng Li, and Zhikai Xu</i>	
Retrospective Labels in Chinese Argumentative Discourses	227
<i>Donghong Liu</i>	
Improve Search by Optimization or Personalization, A Case Study in Sogou Log	231
<i>Jingbin Gao, Muyun Yang, Sheng Li, Tiejun Zhao, and Haoliang Qi</i>	

Machine Translation

Conditional Random Fields for Machine Translation System Combination	237
<i>Tian Xia, Shandian Zhe, and Qun Liu</i>	
A Method of Automatic Translation of Words of Multiple Affixes in Scientific Literature	241
<i>Lei Wang, Baobao Chang, and Janet Harkness</i>	
Hierarchical Pitman-Yor Language Model for Machine Translation	245
<i>Tsuyoshi Okita and Andy Way</i>	
Training MT Model Using Structural SVM	249
<i>Tiansang Du and Baobao Chang</i>	
English-Hindi Automatic Word Alignment with Scarce Resources	253
<i>Eknath Venkataramani and Deepa Gupta</i>	
Sentence Similarity-Based Source Context Modelling in PBSMT	257
<i>Rejwanul Haque, Sudip Kumar Naskar, Andy Way, Marta R. Costa-jussà, and Rafael E. Banchs</i>	
Verb Transfer in a Tamil to Hindi Machine Translation System	261
<i>Sobha Lalitha Devi, Pravin Pralayankar, S. Menaka, T. Bakiyavathi, R. Vijay Sundar Ram, and V. Kavitha</i>	
Lexical Gap in English - Vietnamese Machine Translation: What to Do?	265
<i>Le Manh Hai and Phan Thi Tuoi</i>	
Nominal Transfer from Tamil to Hindi	270
<i>Sobha Lalitha Devi, V. Kavitha, Pravin Pralayankar, S. Menaka, T. Bakiyavathi, and R. Vijay Sundar Ram</i>	

Language Resources

Building Thai FrameNet through a Combination Approach	277
<i>Dhanon Leenoi, Sawittree Jumpathong, and Thepchai Supnithi</i>	
Evaluating the Quality of Web-Mined Bilingual Sentences Using Multiple Linguistic Features	281
<i>Xiaohua Liu and Ming Zhou</i>	

A Semi-Supervised Approach on Using Syntactic Prior Knowledge for Construction Thai Treebank	285
<i>Taneth Ruangrajitpakorn, Prachya Boonkwan, Thepchai Supnithi, and Phiradet Bangcharoensap</i>	
A Proposed Model for Constructing a Yami WordNet	289
<i>Meng-Chien Yang, D. Victoria Rau, and Ann Hui-Huan Chang</i>	
Annotation Guidelines for Hindi-English Word Alignment	293
<i>Rahul Kumar Yadav and Deepa Gupta</i>	
Building Synsets for Indonesian WordNet with Monolingual Lexical Resources	297
<i>Gunawan and Andy Saputra</i>	
A Study of Unique Words in Hawks' Translation of Hong Lou Meng in Comparison with Yang's Translation	301
<i>Yunfang Liang, Lixin Wang, and Dan Yang</i>	
Kazakh Noun Phrase Extraction Based on N-gram and Rules	305
<i>Gulila Altenbek and Ruina Sun</i>	
Spoken Language Processing	
Feature Smoothing and Frame Reduction for Speaker Recognition	311
<i>Santi Nuratch, Panuthat Boonpramuk, and Chai Wutiwiwatchai</i>	
Improved Cantonese Tone Recognition with Approximated F0 Contour: Implications for Cochlear Implants	315
<i>Meng Yuan, Haihong Feng, and Tan Lee</i>	
Combining Sub-bands SNR on Cochlear Model for Voice Activity Detection	319
<i>Qibo Liu, Yi Liu, and Yanjie Li</i>	
Precedence of Emotional Features in Emotional Prosody Processing: Behavioral and ERP Evidence	323
<i>Xuhai Chen and Yufang Yang</i>	
Acoustic Space of Vowels with Different Tones: Case of Thai Language	327
<i>Vaishna Narang, Deepshikha Misra, Ritu Yadav, and Sulaganya Punyayodhin</i>	
A Study of F1 Correlation with F0 in a Tone Language: Case of Thai	330
<i>Sulaganya Punyayodhin, Deepshikha Misra, Ritu Yadav, and Vaishna Narang</i>	
A Contrastive Study of F3 Correlation with F2 and F1 in Thai and Hindi	334
<i>Ritu Yadav, Deepshikha Misra, Sulaganya Punyayodhin, and Vaishna Narang</i>	
Durational Contrast and Centralization of Vowels in Hindi and Thai	339
<i>Deepshikha Misra, Ritu Yadav, Sulaganya Punyayodhin, and Vaishna Narang</i>	
A Study in Comparing Acoustic Space: Korean and Hindi Vowels	343
<i>Hyunkyung Lee and Vaishna Narang</i>	
Author Index	347