2010 International Conference on Asian Language Processing

(IALP 2010)

Harbin, China
28 – 30 December 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
   Biligsaikhan Batjargal, Fuminori Kimura, and Akira Maeda

A Combination of Statistical and Rule-Based Approach for Mongolian Lexical Analysis .................. 7
   Lili Zhao, Jia Men, Congpin Zhang, Qun Liu, Wenbin Jiang, Jinxing Wu,
   and Qing Chang

A Letter Tagging Approach to Uyghur Tokenization ........................................................................11
   Batuer Aisha

Development of Analysis Rules for Bangla Root and Primary Suffix for Universal Networking Language .......................................................... 15
   Md. Nawab Yousuf Ali, Shahid Al Noor, Md. Zakir Hossain, and Jugal Krishna Das

A Suffix-Based Noun and Verb Classifier for an Inflectional Language ............................................ 19
   Navanath Saharia, Utpal Sharma, and Jugal Kalita

Behavior of Word ‘kaa’ in Urdu Language ......................................................................................... 23
   Muhammad Kamran Malik, Aasim Ali, and Shahid Siddiq

Methods to Divide Uyugur Morphemes and Treatments for Exceptions ........................................... 27
   Pu Li and Hao Zhao

Rules for Morphological Analysis of Bangla Verbs for Universal Networking Language .......................... 31
   Md. Nawab Yousuf Ali, Mohammad Zakir Hossain Sarker,
   Ghulam Farooque Ahmed, and Jugal Krishna Das
Discussion on Collation of Tibetan Syllable .................................................................35
Heming Huang and Feipeng Da

A Dictionary Mechanism for Chinese Word Segmentation Based on the Finite Automata ........................................................................................................39
Wu Yang, Liyun Ren, and Rong Tang

Development of Templates for Dictionary Entries of Bangla Roots and Primary Suffixes for Universal Networking Language ............................................43

A Study on "Worry" Separable Words & Its Separable Slots ........................................47
Chunling Li and Xiaoxiao Wang

Syntax and Parsing

Improving Dependency Parsing Using Punctuation .....................................................53
Zhenghua Li, Wanxiang Che, and Ting Liu

A Tree Probability Generation Using VB-EM for Thai PGLR Parser ...........................57
Kanokorn Trakultaweekoon, Taneth Ruangrajitpakorn, Prachya Boonkwan, and Thepchai Supnithi

Research on Verb Subcategorization-Based Syntactic Parsing Postprocess for Chinese Language .................................................................................................61
Jinyong Wang and Xiwu Han

Identification of Maximal-Length Noun Phrases Based on Maximal-Length Preposition Phrases in Chinese ..............................................................................65
Guiping Zhang, Wenjing Lang, Qiaoli Zhou, and Dongfeng Cai

Urdu Noun Phrase Chunking - Hybrid Approach .........................................................69
Shahid Siddiq, Sarmad Hussain, Aasim Ali, Kamran Malik, and Wajid Ali

The Function of Fixed Word Combination in Chinese Chunk Parsing .............................73
Liqun Wang and Shoichi Yokoyama

Problems and Review of Statistical Parsing Language Model .......................................77
Faguo Zhou, Fan Zhang, and Bingru Yang

A General Comparison on Sentences Analysis and Its Teaching Significance between Traditional and Structural Grammars ......................................................81
Jiaying Yu

Semantics

Two Cores in Chinese Negation System: A Corpus-Based View ..................................87
Hio Tong Chan and Chunyu Kit

Finding Semantic Similarity in Vietnamese ..................................................................91
Dat Tien Nguyen and Son Bao Pham

Automatic Metaphor Recognition Based on Semantic Relation Patterns ....................95
Xuri Tang, Weiguang Qu, Xiaohe Chen, and Shiwen Yu
Event Entailment Extraction Based on EM Iteration ................................................................. 101
Zhen Li, Hanjing Li, Mo Yu, Tiejun Zhao, and Sheng Li

On the Semantic Orientation and Computer Identification of the Adverb “Jiù” .................................................................................................................. 105
Lin He and Jiaqin Wu

Semantic Genes and the Formalized Representation of Lexical Meaning ........................................... 110
Dan Hu

Acquisition of Hypernymy-Hyponymy Relation between Nouns for WordNet Building .................................................. 114
Gunawan and Erick Pranata

Algorithm for Conversion of Bangla Sentence to Universal Networking Language ........................................... 118

Construction of the Paradigmatic Semantic Network Based on Cognition ........................................... 122
Xiaofang Ouyang

The Research of Sentence Testing Based on HNC Analysis System of Sentence Category .................................................. 126
Zhiying Liu

Semantic Patterns of Chinese Post-Modified V+N Phrases .................................................. 130
Likun Qiu and Wenxian Zhang

**Information Extraction**

A Grammar-Based Unsupervised Method of Mining Volitive Words .................................................. 137
Jianfeng Zhang, Yu Hong, Yuehui Yang, Jianmin Yao, and Qiaoming Zhu

Using Feature Selection to Speed Up Online SVM Based Spam Filtering ........................................... 142
Yuewu Shen, Guanglu Sun, Haoliang Qi, and Xiaoming He

A Semi-supervised Method for Classification of Semantic Relation between Nominals .................................................. 146
Yuan Chen, Yue Lu, Man Lan, Jian Su, and Zhengyu Niu

XPath-Wrapper Induction for Data Extraction .................................................................................. 150
Nam-Khanh Tran, Kim-Cuong Pham, and Quang-Thuy Ha

A Block Segmentation Based Approach for Web Information Extraction ........................................... 154
Chanwei Wang, Chengjie Sun, Lei Lin, and Xiaolong Wang

Linguistic Features for Named Entity Recognition Using CRFs .................................................................. 158
R. Vijay Sundar Ram, A. Akilandeswari, and Sobha Lalitha Devi

Research on Domain-Adaptive Transfer Learning Method and Its Applications ........................................... 162
Geli Fei and Dequan Zheng
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Theory Based Feature Valuing for Logistic Regression</td>
<td>166</td>
</tr>
<tr>
<td>for Spam Filtering</td>
<td></td>
</tr>
<tr>
<td>Haoliang Qi, Xiaoning He, Yong Han, Muyun Yang, and Sheng Li</td>
<td></td>
</tr>
<tr>
<td>Automatic Named Entity Set Expansion Using Semantic Rules and Wrappers for Unary Relations</td>
<td>170</td>
</tr>
<tr>
<td>Mai-Yu Tran, Tien-Tung Nguyen, Thanh-Son Nguyen, and Hoang-Quynh Le</td>
<td></td>
</tr>
<tr>
<td>Anaphora Resolution of Malay Text: Issues and Proposed Solution Model</td>
<td>174</td>
</tr>
<tr>
<td>Noorhuzaimi Karimah Mohd Noor, Shahrul Azman Noah, Mohd Juzaidin Ab. Aziz, and Mohd Pouzi Hamzah</td>
<td></td>
</tr>
<tr>
<td>Combining Multi-features with Conditional Random Fields for Person Recognition</td>
<td>178</td>
</tr>
<tr>
<td>Suxiang Zhang</td>
<td></td>
</tr>
<tr>
<td>Comparison between Typical Discriminative Learning Model and Generative Model in Chinese Short Messages Service Spam Filtering</td>
<td>182</td>
</tr>
<tr>
<td>Xiaoxia Zheng, Chao Liu, Chengzhe Huang, Yu Zou, and Hongwei Yu</td>
<td></td>
</tr>
<tr>
<td>Chinese Spam Filter Based on Relaxed Online Support Vector Machine</td>
<td>185</td>
</tr>
<tr>
<td>Yong Han, Xiaoning He, Muyun Yang, Haoliang Qi, and Chao Song</td>
<td></td>
</tr>
<tr>
<td>Comment Target Extraction Based on Conditional Random Field &amp; Domain Ontology</td>
<td>189</td>
</tr>
<tr>
<td>Shengchun Ding and Ting Jiang</td>
<td></td>
</tr>
<tr>
<td><strong>Text Understanding and Summarization</strong></td>
<td>195</td>
</tr>
<tr>
<td>Topic-Driven Multi-document Summarization</td>
<td></td>
</tr>
<tr>
<td>Hongling Wang and Guodong Zhou</td>
<td></td>
</tr>
<tr>
<td>Multiple Factors-Based Opinion Retrieval and Coarse-to-Fine Sentiment Classification</td>
<td>199</td>
</tr>
<tr>
<td>Shu Zhang, Wenjie Jia, Yingju Xia, Yao Meng, and Hao Yu</td>
<td></td>
</tr>
<tr>
<td>Chinese Sentence-Level Sentiment Classification Based on Sentiment Morphemes</td>
<td>203</td>
</tr>
<tr>
<td>Xin Wang and Guohong Fu</td>
<td></td>
</tr>
<tr>
<td>Extracting Phrases in Vietnamese Document for Summary Generation</td>
<td>207</td>
</tr>
<tr>
<td>Huong Thanh Le, Rathany Chan Sam, and Phuc Trong Nguyen</td>
<td></td>
</tr>
<tr>
<td>User Interest Analysis with Hidden Topic in News Recommendation System</td>
<td>211</td>
</tr>
<tr>
<td>Mai-Vu Tran, Xuan-Tu Tran, and Huy-Long Uong</td>
<td></td>
</tr>
<tr>
<td>Dependency Tree-Based Anaphoricity Determination for Coreference Resolution</td>
<td>215</td>
</tr>
<tr>
<td>Fang Kong, Jianmei Zhou, Guodong Zhou, and Qiaoming Zhu</td>
<td></td>
</tr>
<tr>
<td>Text Clustering Based on Domain Ontology and Latent Semantic Analysis</td>
<td>219</td>
</tr>
<tr>
<td>Yaxiong Li, Jianqiang Zhang, and Dan Hu</td>
<td></td>
</tr>
</tbody>
</table>
Social Network Mining Based on Wikipedia ................................................................. 223
Fangfang Yang, Zhiming Xu, Sheng Li, and Zhikai Xu

Retrospective Labels in Chinese Argumentative Discourses ........................................ 227
Donghong Liu

Improve Search by Optimization or Personalization, A Case Study in Sogou Log ......................................................................................................................... 231
Jingbin Gao, Muyun Yang, Sheng Li, Tiejun Zhao, and Haoliang Qi

**Machine Translation**

Conditional Random Fields for Machine Translation System Combination .................. 237
Tian Xia, Shandian Zhe, and Qun Liu

A Method of Automatic Translation of Words of Multiple Affixes in Scientific Literature ................................................................................................. 241
Lei Wang, Baobao Chang, and Janet Harkness

Hierarchical Pitman-Yor Language Model for Machine Translation .............................. 245
Tsuyoshi Okita and Andy Way

Training MT Model Using Structural SVM ..................................................................... 249
Tiansang Du and Baobao Chang

English-Hindi Automatic Word Alignment with Scarce Resources ................................ 253
Eknath Venkataramani and Deepa Gupta

Sentence Similarity-Based Source Context Modelling in PBSMT ................................ 257
Rejwanul Haque, Sudip Kumar Naskar, Andy Way, Marta R. Costa-jussà, and Rafael E. Banchs

Verb Transfer in a Tamil to Hindi Machine Translation System ................................... 261
Sobha Lalitha Devi, Pravin Pralayankar, S. Menaka, T. Bakiyavathi, R. Vijay Sundar Ram, and V. Kavitha

Lexical Gap in English - Vietnamese Machine Translation: What to Do? ........................ 265
Le Manh Hai and Phan Thi Tuoi

Nominal Transfer from Tamil to Hindi ............................................................................... 270
Sobha Lalitha Devi, V. Kavitha, Pravin Pralayankar, S. Menaka, T. Bakiyavathi, and R. Vijay Sundar Ram

**Language Resources**

Building Thai FrameNet through a Combination Approach ........................................... 277
Dhanon Leenoi, Sawittree Jumpathong, and Thepchai Supnithi

Evaluating the Quality of Web-Mined Bilingual Sentences Using Multiple Linguistic Features .............................................................................................................. 281
Xiaohua Liu and Ming Zhou
A Semi-Supervised Approach on Using Syntactic Prior Knowledge for Construction Thai Treebank ................................. 285
   Taneth Ruangrajitpakorn, Prachya Boonkwan, Thepchai Supnithi, and Phiradet Bangcharoensap

A Proposed Model for Constructing a Yami WordNet .................................................. 289
   Meng-Chien Yang, D. Victoria Rau, and Ann Hui-Huan Chang

Annotation Guidelines for Hindi-English Word Alignment ........................................ 293
   Rahul Kumar Yadav and Deepa Gupta

Building Synsets for Indonesian WordNet with Monolingual Lexical Resources .................. 297
   Gunawan and Andy Saputra

A Study of Unique Words in Hawks’ Translation of Hong Lou Meng in Comparison with Yang’s Translation ................................................................. 301
   Yunfang Liang, Lixin Wang, and Dan Yang

Kazakh Noun Phrase Extraction Based on N-gram and Rules .................................. 305
   Gulila Altenbek and Ruina Sun

Spoken Language Processing

Feature Smoothing and Frame Reduction for Speaker Recognition ................................. 311
   Santi Nuratch, Panuthat Boompramuk, and Chai Wutiwiwatchai

Improved Cantonese Tone Recognition with Approximated F0 Contour:
Implications for Cochlear Implants ........................................................................ 315
   Meng Yuan, Haihong Feng, and Tan Lee

Combining Sub-bands SNR on Cochlear Model for Voice Activity
Detection .................................................................................................................. 319
   Qibo Liu, Yi Liu, and Yanjie Li

Precedence of Emotional Features in Emotional Prosody Processing:
Behavioral and ERP Evidence ................................................................................ 323
   Xuhai Chen and Yufang Yang

Acoustic Space of Vowels with Different Tones: Case of Thai Language ....................... 327
   Vaishna Narang, Deepshikha Misra, Ritu Yadav, and Sulaganya Punyayodhin

A Study of F1 Correlation with F0 in a Tone Language: Case of Thai ......................... 330
   Sulaganya Punyayodhin, Deepshikha Misra, Ritu Yadav, and Vaishna Narang

A Contrastive Study of F3 Correlation with F2 and F1 in Thai and Hindi .................... 334
   Ritu Yadav, Deepshikha Misra, Sulaganya Punyayodhin, and Vaishna Narang

Durational Contrast and Centralization of Vowels in Hindi and Thai ......................... 339
   Deepshikha Misra, Ritu Yadav, Sulaganya Punyayodhin, and Vaishna Narang

A Study in Comparing Acoustic Space: Korean and Hindi Vowels .............................. 343
   Hyunkyung Lee and Vaishna Narang

Author Index ................................................................................................................... 347