

2011 International Conference on Asian Language Processing

(IALP 2011)

**Penang, Malaysia
15-17 November 2011**



IEEE Catalog Number: CFP1144I-PRT
ISBN: 978-1-4577-1733-8

2011 International Conference on Asian Language Processing

IALP 2011

Table of Contents

Message from the General Chair	x
Message from the Program Chairs	xi
Message from the Local Organizing Chair	xii
Conference Committees	xiii
Program Committee	xiv
Invited Talks	xvi

Phonology, Morphology, Syntax, and Language Model

A Simplified-Traditional Chinese Character Conversion Model Based on Log-Linear Models	3
<i>Yidong Chen, Xiaodong Shi, and Changle Zhou</i>	
Improving Chinese Dependency Parsing with Self-Disambiguating Patterns	7
<i>Likun Qiu, Lei Wu, Kai Zhao, Changjian Hu, and Lingpeng Kong</i>	
Joint Decoding for Chinese Word Segmentation and POS Tagging Using Character-Based and Word-Based Discriminative Models	11
<i>Xinxin Li, Xuan Wang, and Lin Yao</i>	
Natural Language Grammar Induction of Indonesian Language Corpora Using Genetic Algorithm	15
<i>Arya Tandy Hermawan, Gunawan, and Joan Santoso</i>	
Error-Driven Adaptive Language Modeling for Chinese Pinyin-to-Character Conversion	19
<i>Jin Hu Huang and David Powers</i>	
Theoretical Framework of Mongolian Word Segmentation Specification for Information Processing	23
<i>Tong Laga and Xiaobing Zhao</i>	
Research on the Uyghur Information Database for Information Processing	26
<i>Yusup Ebeydulla, Hesengan Abliz, and Azragul Yusup</i>	
Sentence Boundary Detection in Colloquial Arabic Text: A Preliminary Result	30
<i>Afnan A. Al-Subaihini, Hend S. Al-Khalifa, and AbdulMalik S. Al-Salman</i>	
A Study of the Classification and Arrangement Rule of Uygur Morphemes for Information Processing	33
<i>Pu Li and Shuzhen Shi</i>	

Graph-Based Language Model of Long-Distance Dependency	37
<i>Faguo Zhou and Xingang Yu</i>	
BASRAH: Arabic Verses Meters Identification System	41
<i>Zainab A. Khalaf, Maytham Alabbas, and Tien-Ping Tan</i>	
Semantics	
WordNet Editor to Refine Indonesian Language Lexical Database	47
<i>Gunawan, Jessica Felani Wijoyo, I. Ketut Eddy Purnama, and Mochamad Hariadi</i>	
Two Ontological Approaches to Building an Intergrated Semantic Network for Yami <i>ka</i> -Verbs	51
<i>Meng-Chien Yang, Si-Wei Huang, and D. Victoria Rau</i>	
Issues with the Unergative/Unaccusative Classification of the Intransitive Verbs	55
<i>Nitesh Surtani, Khushboo Jha, and Soma Paul</i>	
A Sentence-Level Semantic Annotated Corpus Based on HNC Theory	59
<i>Zhiying Liu, Yaohong Jin, and Chuanjiang Miao</i>	
An Exploration on the Modes of Word Meaning Extension Based on Metaphorical and Metonymic Mechanisms	63
<i>Xiaofang Ouyang</i>	
On the Semantic Orientation and Computer Identification of the Chinese Adverb cai	67
<i>Lin He and Pengbing Chen</i>	
Exploring Both Flat and Structured Features for Number Type Identification of Chinese Personal Noun Phrases	71
<i>Jun Lang</i>	
Discourse	
The Context Imperative Sentences of Modern Chinese	77
<i>Hao Zhao and Kaihong Yang</i>	
Research on Cross-Document Coreference of Chinese Person Name	81
<i>Ji Ni, Fang Kong, Peifeng Li, and Qiaoming Zhu</i>	
Research of Event Pronoun Resolution	85
<i>Ning Zhang, Fang Kong, and Peifeng Li</i>	
Co-reference Resolution in Vietnamese Documents Based on Support Vector Machines	89
<i>Duc-Trong Le, Mai-Vu Tran, Tri-Thanh Nguyen, and Quang-Thuy Ha</i>	
Research of Noun Phrase Coreference Resolution	93
<i>Junwei Gao, Fang Kong, Peifeng Li, and Qiaoming Zhu</i>	
Discourse Structures of English Exposition	97
<i>Donghong Liu and Meizhen Liao</i>	
Linguistic Competency Model for Intentional Agent	101
<i>Sivakumar Ramakrishnan and Vasuky Mohanan</i>	

Text Understanding and Retrieval

The Comparison of Chinese Spam Filter Based on Generative Model and Discriminative Model	107
<i>Yong Han, Yingying Wang, Huafu Ding, and Haoliang Qi</i>	
Applying Grapheme, Word, and Syllable Information for Language Identification in Code Switching Sentences	111
<i>Yin-Lai Yeong and Tien-Ping Tan</i>	
An Integrated Approach Using Conditional Random Fields for Named Entity Recognition and Person Property Extraction in Vietnamese Text	115
<i>Hoang-Quynh Le, Mai-Vu Tran, Nhat-Nam Bui, Nguyen-Cuong Phan, and Quang-Thuy Ha</i>	
A Query Reformulation Model Using Markov Graphic Method	119
<i>Jiali Zuo and Mingwen Wang</i>	
Search Results Clustering Based on a Linear Weighting Method of Similarity	123
<i>Dequan Zheng, Haibo Liu, and Tiejun Zhao</i>	
Extracting Pseudo-Labeled Samples for Sentiment Classification Using Emotion Keywords	127
<i>Sophia Yat Mei Lee, Daming Dai, Shoushan Li, and Kathleen Ahrens</i>	
Imbalanced Sentiment Classification with Multi-strategy Ensemble Learning	131
<i>Zhongqing Wang, Shoushan Li, Guodong Zhou, Peifeng Li, and Qiaoming Zhu</i>	
Formalization and Rules for Recognition of Satirical Irony	135
<i>Lingpeng Kong and Likun Qiu</i>	

Summarization

An Automatic Linguistics Approach for Persian Document Summarization	141
<i>Hossein Kamyar, Mohsen Kahani, Mohsen Kamyar, and Asef Poormasoomi</i>	
Context-Based Persian Multi-document Summarization (Global View)	145
<i>Asef Poormasoomi, Mohsen Kahani, Saeed Varasteh Yazdi, and Hossein Kamyar</i>	
Centroid Integer Selection Model—A High Efficiency Method on Dynamic Multi-document Summarization	150
<i>Meiling Liu, Dequan Zheng, Tiejun Zhao, and Yang Yu</i>	
Corpus Based Extractive Document Summarization for Indic Script	154
<i>P. Vijayapal Reddy, B. Vishnu Vardhan, and A. Govardhan</i>	
Research on Multi-document Summarization Model Based on Dynamic Manifold-Ranking	158
<i>Meiling Liu, Hong Ren, Dequan Zheng, and Tiejun Zhao</i>	

Machine Translation

An Orientation Model for Hierarchical Phrase-Based Translation	165
<i>Xinyan Xiao, Jinsong Su, Yang Liu, Qun Liu, and Shouxun Lin</i>	

Improving Bilingual Lexicon Construction from Chinese-English Comparable Corpora via Dependency Relationship Mapping	169
<i>Hua Xu, Dandan Liu, Longhua Qian, and Guodong Zhou</i>	
A Rule-Based Source-Side Reordering on Phrase Structure Subtrees	173
<i>Fangli Liang, Lei Chen, Miao Li, and Nasun-urtu</i>	
Automatic Acquisition of Chinese-Tibetan Multi-word Equivalent Pair from Bilingual Corpora	177
<i>Minghua Nuo, Huidan Liu, Longlong Ma, Jian Wu, and Zhiming Ding</i>	
Character-Level System Combination: An Empirical Study for English-to-Chinese Spoken Language Translation	181
<i>Jinhua Du</i>	
Mining Parallel Data from Comparable Corpora via Triangulation	185
<i>Thi-Ngoc-Diep Do, Eric Castelli, and Laurent Besacier</i>	
Using Rich Linguistic and Contextual Information for Tree-Based Statistical Machine Translation	189
<i>Bui Thanh Hung, Nguyen Le Minh, and Akira Shimazu</i>	
Research on Element Sub-sentence in Chinese-English Patent Machine Translation	193
<i>Zhiying Liu, Yaohong Jin, and Yuhuan Chi</i>	
Lexical Word Similarity for Re-ranking in Vietnamese-English Named Entity Back Transliteration	197
<i>Thi Hoang Diem Le and Ai Ti Aw</i>	
The Chinese-English Bilingual Sentence Alignment Based on Length	201
<i>Huafu Ding, Lili Quan, and Haoliang Qi</i>	
Optimal Translation Boundaries for BTG-Based Decoding	205
<i>Xiangyu Duan and Min Zhang</i>	
Spoken Language Processing	
Acoustic Space in Motor Disorders of Speech: Two Case Studies	211
<i>Vaishna Narang, Deepshikha Misra, and Garima Dalal</i>	
Adopting Malay Syllable Structure for Syllable Based Speech Synthesizer for Iban and Bidayuh Languages	216
<i>Sarah F. S. Juan, Vyonne Edwin, Chai Yeen Cheong, Jun Choi Lee, and Alvin W. Yeo</i>	
How Vietnamese Attitudes can be Recognized and Confused: Cross-Cultural Perception and Speech Prosody Analysis	220
<i>Dang-Khoa Mac, Eric Castelli, Véronique Aubergé, and Albert Rilliard</i>	
Non-native Accent Pronunciation Modeling in Automatic Speech Recognition	224
<i>Basem H.A. Ahmed and Tien-Ping Tan</i>	
An In-car Chinese Noise Corpus for Speech Recognition	228
<i>Jue Hou, Yi Liu, Chao Zhang, and Shilei Huang</i>	
An Experimental Study on Vietnamese Speech Synthesis	232
<i>Liping Kui, Jian Yang, Bin He, and Enxing Hu</i>	

Development of Acoustic Space in 3 to 5 Years Old Hindi Speaking Children	236
<i>Vaishna Narang, Garima Dalal, and Deepshikha Misra</i>	
Design of a Query-by-Humming System for Hindi Songs Using DDTW Based Approach	240
<i>Prakhar K. Jain, Robin Jain, Hemant A. Patil, and T.K. Basu</i>	
Linear Regression for Prosody Prediction via Convex Optimization	244
<i>Ling Cen, Minghui Dong, and Paul Chan</i>	
Analyzing the Relationship between Formants and Pitch for Singing Voice	248
<i>Hwee Teng Tan and Minghui Dong</i>	
Linguistic Resources and Tools	
Using HTML Tags to Improve Parallel Resources Extraction	255
<i>Yanhui Feng, Yu Hong, Wei Tang, Jianmin Yao, and Qiaoming Zhu</i>	
Automatic Labeling and Phonetic Assessment for an Unknown Asian Language: The Case of the “Mo Piu” North Vietnamese Minority (early results)	260
<i>Geneviève Caelen-Haumont, Sethserey Sam, and Eric Castelli</i>	
Automatic Construction of Chinese-Mongolian Parallel Corpora from the Web Based on the New Heuristic Information	264
<i>Zede Zhu, Miao Li, Lei Chen, and Shouguo Zheng</i>	
Developing Bengali Speech Corpus for Phone Recognizer Using Optimum Text Selection Technique	268
<i>Sandipan Mandal, Biswajit Das, Pabitra Mitra, and Anupam Basu</i>	
Polarity Shifting: Corpus Construction and Analysis	272
<i>Xiaoqian Zhang, Shoushan Li, Guodong Zhou, and Hongxia Zhao</i>	
Building a Rule-Based Malay Text Segmentation Tool	276
<i>Bali Ranaivo-Malançon</i>	
Language Learning	
The Phoneme-Level Articulator Dynamics for Pronunciation Animation	283
<i>Sheng Li, Lan Wang, and En Qi</i>	
The Effect of Arabic Language on Reading English for Arab EFL Learners: An Eye Tracking Study	287
<i>Kholod S. Al-Khalifah and Hend S. Al-Khalifa</i>	
A Study on Eye Movement of Thai Students Reading Chinese Texts with or without Marks for Word Boundaries	291
<i>Yumei Jiao and Peng Yu</i>	
Games for Academic Vocabulary Learning through a Virtual Environment	295
<i>Kiran Pala, Anil Kumar Singh, and Suryakanth V. Gangashetty</i>	
Author Index	299