

2013 International Conference on Asian Language Processing

(IALP 2013)

**Urumqi, China
17-19 August 2013**



**IEEE Catalog Number: CFP1344I-POD
ISBN: 978-1-4799-1096-0**

2013 International Conference on Asian Language Processing

IALP 2013

Table of Contents

Message from General Chairsxi
Message from Program Chairsxii
Committeesxiii
Keynotesxvii
Organizers and Sponsorsxxi

Discourse Analysis

Recognizing Chinese Elementary Discourse Unit on Comma	3
<i>Shengqin Xu and Peifeng Li</i>	
Judgment, Extraction and Selective Restriction of Chinese Eventive Verb	7
<i>Mengxiang Wang, Houfeng Wang, and Longkai Zhang</i>	
Topic and Its Negation in Chinese Sentences	11
<i>Lin He and Qiong Peng</i>	
Discourse Topic in Anaphora Resolution and Discourse Construction	15
<i>Donghong Liu</i>	

Information Extraction and Retrieval

An Empirical Evaluation of Dimensionality Reduction Using Latent Semantic Analysis on Hindi Text	21
<i>Karthik Krishnamurthi, Ravi Kumar Sudi, Vijayapal Reddy Panuganti, and Vishnu Vardhan Bulusu</i>	
Relationship Extraction Tactics of Chinese Entity Based on Formal Concept Connectivity Distance	25
<i>Chunlei Cheng</i>	
Mining Recipes in Microblog	29
<i>Shengyu Liu, Qingcai Chen, Shanshan Guan, Xiaolong Wang, and Huimiao Shi</i>	
Combination of Unsupervised Keyphrase Extraction Algorithms	33
<i>Zede Zhu, Miao Li, Lei Chen, Zhenxin Yang, and Sheng Chen</i>	

Information Retrieval Model Combining Sentence Level Retrieval	37
<i>Jiali Zuo, Mingwen Wang, Jianyi Wan, and Wenbing Luo</i>	

Linguistics and Language Study

Research on Prosody Features of Mongolian Traditional Folk Long Song	43
<i>Yasheng Jin and Wenmin Liu</i>	
Categorization and Identification of Fragments with Shi Plus Punctuation	47
<i>Guonian Wang and Lin He</i>	
On the Formation and Semantic Meaning of the Title [[V+O]+O] in Mandarin	51
<i>Hongwu Xue</i>	
NLP-Oriented Study on the Imperative Sentence with Interrogative Mood	54
<i>Pu Li and Hao Zhao</i>	
A Cognitive Schema for Transferred Meanings of the Body-Disease Lexicon in Chinese	57
<i>Jia Li</i>	
Research of Modern Uyghur Word Frequency Statistical Technology	60
<i>Azragul, Nianmei, and Yasen Yimin</i>	

Machine Translation

New Language Resources for Arabic: Corpus Containing More Than Two Million Words and a Corpus Processing Tool	67
<i>Abdulmohsen Al-Thubaity, Marwa Khan, Manal Al-Mazrua, and Maram Al-Mousa</i>	
Research and Implementation of the Uyghur-Chinese Personal Name Transliteration Based on Syllabification	71
<i>Alim Murat, Azragul Yusup, and Yusup Abaydulla</i>	
Rule Refinement for Spoken Language Translation by Retrieving the Missing Translation of Content Words	75
<i>Linfeng Song, Jun Xie, Xing Wang, Yajuan Lv, and Qun Liu</i>	
Transfer Grammar in Tamil-Hindi MT System	79
<i>Sobha Lalitha Devi, Sindhuja Gopalan, and Vijay Sundar Ram</i>	
Analysis and Evaluation of Terminology Translation Consistency in Scientific and Technical Literature	83
<i>Baosheng Yin, Xiaodong Yue, Dongfeng Cai, and Guiping Zhang</i>	
Implementation of Chinese-Uyghur Bilateral EBMT System	87
<i>Kahaerjiang Abiderexiti, Tianfang Yao, Tuergen Yibulayin, Aishan Wumaier, and Yasen Yiming</i>	
Improvements in Statistical Phrase-Based Interactive Machine Translation	91
<i>Dongfeng Cai, Hua Zhang, and Na Ye</i>	

Findings and Considerations in Active Learning Based Framework for Resource-Poor SMT	95
<i>Jinhua Du and Meng Zhang</i>	
Mining Parallel Corpus from Sina Microblog	99
<i>Haitao Xing, Muyun Yang, Haoliang Qi, Sheng Li, and Tiejun Zhao</i>	
Machine Translation Approach for Vietnamese Diacritic Restoration	103
<i>Thi Ngoc Diep Do, Duy Binh Nguyen, Dang Khoa Mac, and Do Dat Tran</i>	
A Purely Monotonic Approach to Machine Translation for Similar Languages	107
<i>Ye Kyaw Thu, Andrew Finch, Eiichiro Sumita, and Yoshinori Sagisaka</i>	

Sentiment Analysis

Subjectivity Classification of Filipino Text with Features Based on Term Frequency—Inverse Document Frequency	113
<i>Ralph Vincent J. Regalado, Jenina L. Chua, Justin L. Co, and Thomas James Z. Tiam-Lee</i>	
Mining Semantic Orientation of Multiword Expression from Chinese Microblogging with Discriminative Latent Model	117
<i>Xiao Sun, Chengcheng Li, Chenyi Tang, and Fuji Ren</i>	
Exploiting Hierarchical Discourse Structure for Review Sentiment Analysis	121
<i>Fei Wang and Yunfang Wu</i>	
Data Quality Controlling for Cross-Lingual Sentiment Classification	125
<i>Shoushan Li, Yunxia Xue, Zhongqing Wang, Sophia Yat Mei Lee, and Chu-Ren Huang</i>	
Sentiment Classification with Polarity Shifting Detection	129
<i>Shoushan Li, Zhongqing Wang, Sophia Yat Mei Lee, and Chu-Ren Huang</i>	

Spoken Language Processing

Tonal Effects on Vowel Production in Standard Chinese—An Electropalatographic and Acoustic Study	135
<i>Jinghua Zhang, Xiaosheng Pan, and Yinghao Li</i>	
A Novel Gaussian Filter-Based Automatic Labeling of Speech Data for TTS System in Gujarati Language	139
<i>Swati Talesara, Hemant A. Patil, Tanvina Patel, Hardik Sailor, and Nirmesh Shah</i>	
Use of PLP Cepstral Features for Phonetic Segmentation	143
<i>Bhavik B. Vachhani and Hemant A. Patil</i>	
Speech Summarization without Lexical Features for Mandarin Presentation Speech	147
<i>Jian Zhang and Huaqiang Yuan</i>	

An Improved MMSE-LSA Speech Enhancement Algorithm Based on Human Auditory Masking Property	151
<i>Yong Zhang and Yi Liu</i>	
Application of Tucker Decomposition in Speech Signal Feature Extraction	155
<i>Lidong Yang, Jing Wang, Xiang Xie, and Jingming Kuang</i>	
The Acoustic Research of Stops CV Structure Coarticulation in Amdo Tibetan Xiahe Dialect	159
<i>Shiliang Lv, Yasheng Jin, Ning Ma, and Xuechen Yin</i>	
Speech Recognition Research on Uyghur Accent Spoken Language	163
<i>Yating Yang, Bo Ma, Xinyu Tang, and Osman Turghun</i>	
Improving the Accuracy of Large Vocabulary Continuous Speech Recognizer Using Dependency Parse Tree and Chomsky Hierarchy in Lattice Rescoring	167
<i>Kai Sze Hong, Tien-Ping Tan, and Enya Kong Tang</i>	
Varying or Unvarying-Logarithmic Quotient Model of Vowel Formants	171
<i>Xuewen Zhou</i>	

Syntactic and Semantic Analysis

Improving Chinese Parsing with Special-Case Probability Re-estimation	177
<i>Yu-Ming Hsieh, Su-Chu Lin, Jason S. Chang, and Keh-Jiann Chen</i>	
Dependency Parsing for Traditional Mongolian	181
<i>Xiangdong Su, Guanglai Gao, and Xueliang Yan</i>	
The Annotation Scheme for Uyghur Dependency Treebank	185
<i>Samat Mamitimin, Turgun Ibrahim, and Marhaba Eli</i>	
A Tibetan Ontology Concept Acquisition Method Based on HowNet and Chinese-Tibetan Dictionary	189
<i>Xinmin Jiang and Lirong Qiu</i>	
Evaluating the Difficulty of Concepts on Domain Knowledge Using Latent Semantic Analysis	193
<i>Tao-Hsing Chang, Yao-Ting Sung, and Yao-Tung Lee</i>	
Pronominal Resolution in Tamil Using Tree CRFs	197
<i>R. Vijay Sundar Ram and Sobha Lalitha Devi</i>	
The Frame Design of Mongolian Noun Semantic Network	201
<i>Hasi and Enbo Tang</i>	

Text Classification and Clustering

Combination of ROSVM and LR for Spam Filter	207
<i>Yadong Wang, Haoliang Qi, Hong Deng, and Yong Han</i>	
A Method for Network Topic Attention Forecast Based on Feature Words	211
<i>Chunlei Yan, Shumin Shi, Heyan Huang, and Ruijing Li</i>	

Broadcast News Story Clustering via Term and Sentence Matching	215
<i>Foong Kuin Yow and Tien-Ping Tan</i>	
Tibetan Text Classification Based on the Feature of Position Weight	220
<i>Hui Cao and Huiqiang Jia</i>	

Word Segmentation and POS Tagging

A Unified Model for Joint Chinese Word Segmentation and POS Tagging with Heterogeneous Annotation Corpora	227
<i>Jiayi Zhao, Xipeng Qiu, and Xuanjing Huang</i>	
Feature Abstraction for Lightweight and Accurate Chinese Word Segmentation	231
<i>Le Tian, Xipeng Qiu, and Xuanjing Huang</i>	
Uyghur Stem-Suffix Segmentation and POS-Tagging Based on Functional Suffixes	235
<i>Haibo Wang, Yiqing Zu, Lang Wu, and Litip Tohti</i>	
Tibetan Word Segmentation Based on Word-Position Tagging	239
<i>Caijun Kang, Di Jiang, and Congjun Long</i>	
The Comparative Research on the Segmentation Strategies of Tibetan Bounded-Variant Forms	243
<i>Congjun Long, Caijun Kang, and Di Jiang</i>	

Others

A Computer-Assist Algorithm to Detect Repetitive Stuttering Automatically	249
<i>Junbo Zhang, Bin Dong, and Yonghong Yan</i>	
An Efficient Algorithm of Chinese String Sort in User-Defined Sequence	253
<i>Haijun Zhang and Shumin Shi</i>	
On Teaching Strategies of Light Tone in Chinese to Foreign Students	257
<i>Mingyue Wang</i>	
A Tentative Study on Language Model Based Solution to Multiple Choice of CET-4	261
<i>Zhihang Fan, Muyun Yang, Tiejun Zhao, and Sheng Li</i>	
Multi-thread Multi-keywords Matching Approach for Uyghur Text	265
<i>Xinyuan Zhao and Adili</i>	
Using Mutual Information Criterion to Design an Effective Lexicon for Chinese Pinyin-to-Character Conversion	269
<i>Wei Li, Jinsong Zhang, Yanlu Xie, Xiaoyun Wang, Masafumi Nishida, and Seiichi Yamamoto</i>	
A Study on Japanese Students (Intermediate Level) Reading Chinese Texts With or Without Marks for Word Boundaries	273
<i>Yu Peng</i>	

Research of the Modern Uyghur Data Analysis Technology278
Mengchen Pan, Xiangwei Qi, and Weimin Pan

Author Index282