2016 International Conference on Asian Language Processing (IALP 2016)

Tainan, Taiwan
21-23 November 2016

IEEE Catalog Number: CFP1644I-POD
ISBN: 978-1-5090-0923-7

Copyright © 2016 by the Institute of Electrical and Electronics Engineers, Inc
All Rights Reserved

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number: CFP1644I-POD
ISBN (Print-On-Demand): 978-1-5090-0923-7
ISBN (Online): 978-1-5090-0922-0

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

Technical Papers

Session	Speech/Spoken Language Processing
Date / Time	November 21, 2016 / 11:00 – 12:00
Venue	Lecture Room 1
Chair	Dr. Yanfeng Lu

Comparison on Neural Network Based Acoustic Model in Mongolian Speech Recognition 1
Hongwei Zhang, Feilong Bao, Guanglai Gao and Hui Zhang

Dialog State Tracking and Action Selection Using Deep Learning Mechanism for Interview Coaching 6
Ming-Hsiang Su, Kun-Yi Huang, Tsung-Hsien Yang, Kuan-Jung Lai and Chung-Hsien Wu

Policy Optimization of Dialogue Management in Spoken Dialogue System for Out-of-Domain Utterances 10
Yuhong Xu, Peijie Huang, Jiecong Tang, Qiangjia Huang, Zhenpeng Deng, Weimou Peng and Jiajie Lu

Dialogue Act Recognition for Chinese Out-of-Domain Utterances using Hybrid CNN-RF 14
Jundong Wang, Peijie Huang, Qiangjia Huang, Zixuan Ke and Piyuan Lin

Session	Language Resources
Date / Time	November 21, 2016 / 11:00 – 12:00
Venue	Lecture Room 2
Chair	Prof. Baoli Li

Annotating Chinese Noun Phrases Based on Semantic Dependency Graph 18
Yimeng Li and Yanqiu Shao

Sense Annotated Hindi Corpus 21
Satyendr Singh and Tanveer J. Siddiqui

Towards Building a Standard Dataset for Arabic Keyphrase Extraction Evaluation 26
Muhammad Helmy, Marco Basaldella, Eddy Maddalena, Stefano Mizzaro and Gianluca Demartini

Semantic Annotation for Mandarin Verbal Lexicon 30
Mei-chun Liu and Jui-ching Chang

Session	NLP Applications
Date / Time	November 21, 2016 / 14:00 – 15:15
Venue	Lecture Room 1
Chair	Dr. Van Hai Do

Information Extraction and Text Mining of Ancient Vattezhuthu Characters in Historical Documents Using Image Zoning 37
E.K. Vellingiriraj, M. Balamurugan and P. Balasubramanie

Content-based Approach for Vietnamese Spam SMS Filtering 41
Thai-Hoang Pham and Phuong Le-Hong

Detecting Representative Web Articles Using Heterogeneous Graphs 45
Richeng Xuan and Sang-goo Lee

Word Clustering for Parallelism in Classical Chinese Poems 49
John Lee and Mengqi Luo

Annotation Scheme for Legal Discourse Information and Hierarchical Levels 53
Hong Wang and Yunfeng Ge

Session	Machine Learning for Natural Language (I)
Date / Time	November 21, 2016 / 14:00 – 15:15
Venue	Lecture Room 2
Chair	Prof. Chia-Ping Chen

Syntactic Characteristics and Similarities of Japanese Authors' Writing Styles:A Kernel-Based Approach 59
Eriko Kanagawa and Takeshi Okadome

An Initial Study of Indonesian Semantic Role Labeling and Its Application on Event Extraction 63
Ade Romadhony, Ayu Purwarianti and Lisa Madlberger

Generating Manipuri English Pronunciation Dictionary Using Sequence Labelling Problem 67
Rajlakshmi Saikia and Sanasam Ranbir Singh

Recurrent Neural Network-based Language Models with Variation in Net Topology, Language, and Granularity 71
Tzu-Hsuan Yang, Tzu-Hsuan Tseng and Chia-Ping Chen

Verifying the Long-range Dependency of RNN Language Models 75
Tzu-Hsuan Tseng, Tzu-Hsuan Yang and Chia-Ping Chen

Session	Machine Translation and Word Sense Disambiguation
Date / Time	November 21, 2016 / 15:45 – 17:00
Venue	Lecture Room 1
Chair	Dr. Wei-Yun Ma

The Effect of Shallow Segmentation for English-Tigrinya Statistical Machine Translation 79
Yemane Tedla and Kazuhide Yamamoto

Proper Noun Recognition in Cross-Language Record Linkage by Exploiting Transliterated Words 83
Yuting Song, Taisuke Kimura, Biligsaikhan Batjargal and Akira Maeda

Representing Action Structure through Semantic Roles of Physiotherapy Instructions 87
Sandeep Kumar Dash, Partha Pakray and Alexander Gelbukh

Word Sense Implantation by Orthographical Conversion 91
Kazuhide Yamamoto and Yuki Mikami

A Corpus-based Analysis of Filipino Writing Errors 95
Manolito Octaviano, Matthew Philip Go, Allan Borra and Nathaniel Oco

Session	Under-resourced Language Studies
Date / Time	November 21, 2016 / 15:45 – 17:00
Venue	Lecture Room 2
Chair	Dr. John Lee

Gotong Royong in NLP Research - A Mobile Tool for Collaborative Text Annotation in Indonesia 99
Lisa Madlberger, Ade Romadhony and Ayu Purwarianti

Annotation Schemes for Constructing Uyghur Named Entity Relation Corpus 103
Kahaerjiang Abiderexiti, Maihemuti Maimaiti, Tuergen Yibulayin and Aishan Wumaier

Construction of a Bilingual Annotated Corpus with Chinese Buddhist Translation and their Sanskrit Parallels 108
Wei Huangfu, Qingzhi Zhu and Bing Qiu

Speech recognition of under-resourced languages using mismatched transcriptions 112
Van Hai Do, Nancy F. Chen, Boon Pang Lim and Mark Hasegawa-Johnson

A System for Extracting F0 Contours of Lexical Tones Using Adaptive IIR Notch Filter with Harmonic Suppression 116
Siripong Potisuk

Session	Sentiment Analysis and Question Answering
Date / Time	November 22, 2016 / 11:00 – 12:00
Venue	Lecture Room 1
Chair	Dr. Chao-Hong Liu

A Regression Approach to Valence-Arousal Ratings of Words from Word Embedding 120
Minglei Li, Yunfei Long and Qin Lu

Extracting Explanatory Segments from Chinese Product Reviews via Semantic Pattern Matching 124
Liang Zhang, Yu He and Guohong Fu

History Question Classification and Representation for Chinese Gaokao 129
Ke Yu, Qiuzhi Liu, Yuqing Zheng, Tiejun Zhao and Dequan Zheng

Japanese Orthographical Normalization Does Not Work for Statistical Machine Translation 133
Kazuhide Yamamoto and Kanji Takahashi

Session	Linguistics
Date / Time	November 22, 2016 / 11:00 – 12:00
Venue	Lecture Room 2
Chair	Dr. Hong Wang

A Unified Representation of Attributes and Their Semantic Composition 137
Su-Chu Lin, Wei-Yun Ma and Yueh-Yin Shih

The Thematic and Syntactic Analysis of yu 于 Constructions in Archaic Chinese 141
Mengbin Liu

Is the Movement Mentally Simulated in Processing Fictive Motion Sentences in Mandarin? 146
Shu-Ping Gong and Zhao-Ying Huang

A Preliminary Study of Chinese Hierarchical Grammar Development 150
Jia-Fei Hong and Zi-Yin Chen

Date / Time	November 22, 2016 / 13:00 – 14:00
Venue	Lecture Room 1

Overview of the IALP 2016 Shared Task on Dimensional Sentiment Analysis for Chinese Words 156
Liang-Chih Yu, Lung-Hao Lee and Kam-Fai Wong

Aicyber's System for IALP 2016 Shared Task:Character-enhanced Word Vectors and Boosted Neural Networks 161
Steven Du and Xi Zhang

CKIP Valence-Arousal Predictor for IALP 2016 Shared Task 164
Hsin-Yang Wang and Wei-Yun Ma

Evaluation of Weighted Graph and Neural Network Models on Predicting the Valence-Arousal Ratings of Chinese Words 168
Wei-Chieh Chou, Chin-Kui Lin, Yuan-Fu Liao and Yih-Ru Wang

Session	Poster Session
Date / Time	November 22, 2016 / 14:00 – 15:20
Venue	Multi-function Room

A Method of Automatic Recognition of Attributive Clauses in Chinese Language 172
Lei Wang, Houfeng Wang, Weiguang Qu and Shiwen Yu

Design and Realization of a Graphical Editing System for Mongolian Dependency Treebank 176
S. Loglo

Automatic Discrimination System for Chinese Labial Aspirated and Unaspirated Sounds Pronounced by Japanese Students 180
Akemi Hoshino

Automatic Extraction and Recommendation of Grammar Points in L2 Chinese 184
Tianbao Song, Jing He, Weiming Peng and Jihua Song

The Communication Potential Estimation of Chinese Borrowings in English 189
Rui Xin and Xiaolan Lei

Graph-Based Short Text Entity Linking:A Data Integration Perspective 193
Bo Ma, Yating Yang, Xi Zhou and Lei Wang

Application of Imperative Strategies 198
Pu Li

The Usage of haodai 202
Shuzhen Shi and Shuqin Shi

A Novel Decoding Framework for Extractive Speech Summarization with Rhetorical Structure Modeling 206
Jian Zhang and Huaqiang Yuan

Design and Implementation of Online Learning Software of Mongolian Idioms 210
Hai Yinhua

The Evaluation Standards of Semantic Roles:Based on Semantic Dependency Graph Bank 214
Xinghui Cheng and Yanqiu Shao

Research on Question Classification for Automatic Question Answering 218
Shihua Xu, Gang Cheng and Fang Kong

Social Circle Detection Based on Micro-blogging Topic and User Follow Relationship 222
Jiaying Hou, Zhengtao Yu, Xudong Hong and Liren Wang

A Semi-automatic Approach to Identifying and Unifying Ambiguously Encoded Arabic-Based Characters 228
Sardar Jaf

Hierarchical Pre-reordering model for Patent Machine Translation 232
Renfen Hu, Kai Zhao, Hongzheng Li, Yun Zhu and Yaohong Jin

A Hybrid Strategy Method For Patent Machine Translation 236
Zhiying Liu and Kai Zhao

Tonal and Non-Tonal Intonation in Yichang dialect 240
Yan Li

Fundamental Tools and Resource Are Available for Vietnamese Analysis 246
Kanji Takahashi and Kazuhide Yamamoto

Mongolian Prosodic Phrase Prediction Using Suffixes Segmentation 250
Rui Liu, Feilong Bao, Guanglai Gao and Weihua Wang

Developing Learner Corpus Annotation for Chinese Grammatical Errors 254
Lung-Hao Lee, Li-Ping Chang and Yuen-Hsien Tseng

Chinese Word Segmentation based on Conditional Random Fields with Character Clustering 258
Liping Du, Xiaoge Li, Chunli Liu, Rui Liu, Xian Fan, Jianing Yang, Dayi Lin and Mian Wei

Improved Arabic Characters Recognition by Combining Multiple Machine Learning Classifiers 262
Maytham Alabbas, Raidah Khaudeyer and Sardar Jaf

Construction of the Basic Sentence-pattern Instance Database Based on the International Chinese Textbook Treebank 266
Shuqin Zhu, Yinxia Zhang, Weiming Peng and Jihua Song

A Machine Learning Approach for Authorship Attribution for Bengali Blogs 271
Shanta Phani, Shibamouli Lahiri and Arindam Biswas

A Syntactic Analysis on the classifier-noun construction in Chinese 275
Fengcun An and Lei Zhao

A Study on the Construction of a Grade-Level Reading Corpus for TCSL 279
Juan Li, Lijiao Yang, Liang Wen, Zhiying Liu and Yongjie Lu

A Method for Word Sense Disambiguation Combining Contextual Semantic Features 283
Liang Wen, Juan Li, Yaohong Jin and Yongjie Lu

Web Content Extraction Based on Maximum Continuous Sum of Text Density 288
Kai Sun, Miao Li, Jinhua Du, Lei Chen, Zhengxin Yang, Yi Gao and Sha Fu

Valence-Arousal Ratings Prediction with Co-occurrence Word-embedding 293
Yan Gao, Jing Xu, Xu Yang and Bin Xu

IASL Valence-Arousal Analysis System at IALP 2016 shared Task:Dimensional Sentiment Analysis for Chinese Words 297
Yu-Lun Hsieh, Chen-Ann Wang, Ying-Wei Wu, Yung-Chun Chang and Wen-Lian Hsu

Dimensional Sentiment Analysis of Traditional Chinese Words Using Pre-trained Not-quite-right Sentiment Word Vectors and Supervised Ensemble Models 300
Feixiang Wang, Yunxiao Zhou and Man Lan

Valence-Arousal Prediction of Chinese Words with Multi-layer Corpora 304
Xinrui Zhang, Piyuan Lin, Siyuan Chen, Hongjie Cen, Jundong Wang, Qiangjia Huang, Yuhong Xu, Jiecong Tang and Peijie Huang

Constructing a Chinese Valence-Arousal Dictionary with Multiple Heterogeneous Lexicons 308
Tao-Hsing Chang and Yi-Ting Siao

Dimensional Sentiment Analysis for Chinese Words Based on Synonym Lexicon and Word Embedding 312
Wei Cheng, Yuansheng Song, Yue Zhu and Ping Jian

Valence-Arousal Ratings Prediction of Chinese Words Using Similarity Measures Based on Word2Vec 317
Chao-Hong Liu, Qun Liu and Ching-Hsien Lee

A Comparison of Approaches to Affective Rating of Chinese Words on Valence-Arousal Space 320
Bo Peng, Jinnan Yang, Xutao Yang, Dan Xu and Xuejie Zhang

Learning Dimensional Sentiment of Traditional Chinese Words with Word Embedding and Support Vector Regressions 324
Baoli Li

Dimensional Sentiment Analysis in Valence-Arousal for Chinese Words by Linear Regression 328
Jui-Feng Yeh, Tai-You Kuang, Yu-Jui Huang and Mei-Rong Wu

Dimentional Sentiment Analysis by Synsets and Sense Definitions 332
Chuan-Jie Lin and Hao-Tsung Chang

Session	Text Mining
Date / Time	November 22, 2016 / 15:50 – 16:50
Venue	Lecture Room 1
Chair	Prof. Kazuhide Yamamoto

Improving the Effectiveness of POI Search by Associated Information Summarization 336
Hsiu-Min Chuang, Chia-Hui Chang and Chung-Ting Cheng

Leveraging Arabic Morphology and Syntax for Achieving Better Keyphrase Extraction 340
Muhammad Helmy, Dario De Nart, Dante Degl'Innocenti and Carlo Tasso

Importance Weighted Feature Selection Strategy for Text Classification 344
Baoli Li

The Extraction and Application of Constitutive Role of Nouns in Chinese 348
Mengxiang Wang, Houfeng Wang and Luexing Zheng

Session	Machine Learning for Natural Language (II)
Date / Time	November 22, 2016 / 15:50 – 16:50
Venue	Lecture Room 2
Chair	Prof. Guohong Fu

Unsupervised Ensemble Learning for Vietnamese Multisyllabic Word Extraction 353
Wuying Liu and Lin Wang

Named Entity Recognition on Indonesian Microblog Messages 358
Natanael Taufik, Alfan F. Wicaksono and Mirna Adriani

Using Topic Modelling to Make Sense of Typhoon-related Tweets 362
Cerino Ligutom III, Jay Vincent Orio, Dyannah Alexa Marie Ramacho, Chuchi Montenegro, Rachel Edita Roxas and Nathaniel Oco

Modeling Morpheme Triplets with a Three-level Hierarhical Dirichlet Process 366
Serkan Kumyol and Burcu Can