

# **2017 International Conference on Asian Language Processing (IALP 2017)**

**Singapore  
5 – 7 December 2017**



**IEEE Catalog Number: CFP1744I-POD**  
**ISBN: 978-1-5386-1982-7**

**Copyright © 2017 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:	CFP1744I-POD
ISBN (Print-On-Demand):	978-1-5386-1982-7
ISBN (Online):	978-1-5386-1981-0

**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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

CURRAN ASSOCIATES INC.  
**proceedings**  
.com

## Technical Papers

<b>Session 1</b>	Machine learning for natural language I
<b>Date/Time</b>	Tuesday, 05 December 2017 / 10:30 am – 12:00 pm
<b>Chair</b>	Prof. <b>Qi Su</b> , <i>Peking University, China</i>

- 023 Joint Extraction of Argument Components and Relations.....1**  
*Yang Du, Minglan Li and Mengxue Li*
- 041 Joint Learning of Contextal and Global Features for Named Entity Disambiguation.....5**  
*Bo Ma, Tonghai Jiang, Yating Yang, Xi Zhou and Lei Wang*
- 058 Mining Tibetan-Chinese Bilingual Entities from Wikipedia.....9**  
*Tao Jiang, Hongzhi Yu, Xiangzhen He and Xianghe Meng*
- 094 Dynamic topic mining for Microblog fused with user's behavior and time window.....13**  
*Fei Wu, Zhuo Wang, Zhengtao Yu, Liren Wang and Feng Zhou*
- 113 Corpus for the Legal Information Processing System (CLIPS): a Chinese legal corpus annotated with discourse information.....18**  
*Hong Wang and Yunfeng Ge*
- 148 Towards Improving the Performance of Chat Oriented Dialogue System.....23**  
*Ridong Jiang and Rafael E. Banchs*

<b>Session 2</b>	Speech recognition and synthesis
<b>Date/Time</b>	Tuesday, 05 December 2017 / 10:30 am – 12:00 pm
<b>Chair</b>	Dr. <b>Yanfeng Lu</b> , <i>Institute for Infocomm Research, Singapore</i>

- 102 Investigating Multi-task Learning for Automatic Speech Recognition with Code-switching between Mandarin and English.....27**  
*Xiao Song, Yuexian Zou, Shilei Huang, Shaobin Chen and Yi Liu*
- 024 Isolated Digit Filipino Speech Recognition through Spectrogram Image Classification: Towards Application in a Disaster Preparedness Participatory Toolkit.....31**  
*Julie Ann Salido, Nathaniel Oco, Rachel Roxas, Emmanuel Malaay, Michael Simora and Ronald John Cabatic*
- 052 Transfer learning for children's speech recognition.....36**  
*Rong Tong, Lei Wang and Bin Ma*
- 071 Multimodal Learning using 3D Audio-Visual Data for Audio-Visual Speech Recognition.....40**  
*Rongfeng Su, Lan Wang and Xunying Liu*
- 090 On the analysis and evaluation of prosody conversion techniques.....44**  
*Berrak Sisman, Grandee Lee, Haizhou Li and Kay Chen Tan*
- 118 On Some Problems about the Text in Mongolian Speech Synthesis.....48**  
*Qi Bailing*

<b>Session 3</b>	Language learning
<b>Date/Time</b>	Tuesday, 05 December 2017 / 10:30 am – 12:00 pm
<b>Chair</b>	Prof. <b>Yanping Li</b> , <i>Nanjing University of Posts &amp; Telecommunication, China</i>

- 022 Improving Pronunciation Erroneous Tendency Detection with Convolutional Long Short-Term Memory.....52**  
*Longfei Yang, Yingming Gao, Yanlu Xie and Jinsong Zhang*
- 057 Correcting Misuse of Japanese Visually Similar Characters.....57**  
*Youichiro Ogawa and Kazuhide Yamamoto*
- 059 Experimental Research of Mandarin Diphthongs Produced by Uyghur Learners.....61**  
*Yultuz Rapkat, Gulnur Arkin and Askar Hamdulla*
- 079 Chinese teaching material readability assessment with contextual information.....66**  
*Hao Liu, Si Li, Jianbo Zhao, Zuyi Bao and Xiaopeng Bai*
- 119 A New Exploration of Diagrammatic Treebank in International Chinese Teaching.....70**  
*Yinbing Zhang, Jihua Song, Weiming Peng, Dongdong Guo and Canran Sun*
- 125 Evaluating the Intended Learning Outcome of Ancient Chinese Teaching Materials with Teaching-oriented Corpus.....75**  
*Bing Qiu*

<b>Session 4</b>	Question answering
<b>Date/Time</b>	Tuesday, 05 December 2017 / 01:30 pm – 03:15 pm
<b>Chair</b>	Dr. <b>Shuangyong Song</b> , <i>Alibaba Group</i>

- 012 Information Entropy-Informed Sentence Representation for Question Classification.....79**  
*Jin Gao, Miao Li, Lei Chen, Jinhua Du and Rongqiang Ma*
- 014 Intension Classification of User Queries in Intelligent Customer Service System.....83**  
*Shuangyong Song, Haiqing Chen and Zhiwei Shi*
- 048 Extracting Disease-Symptom Relationships from Health Question and Answer Forum.....87**  
*Christian Halim, Alfian F. Wicaksono and Mirna Adriani*
- 076 Combine Multi-features with Deep Learning for Answer Selection.....91**  
*Yuqing Zheng, Chenghe Zhang, Dequan Zheng and Feng Yu*
- 085 Domain Independent Keyword Identification for Question Answering.....95**  
*Prathyusha Jwalapuram and Radhika Mamidi*
- 107 Towards a Deep Learning Powered Query Engine for Urban Planning.....99**  
*Yon Shin Teo, Zihong Yuan, Yangfan Zhang, Valerie Phang and Wee Siong Ng*
- 130 Single Turn Chinese Emotional Conversation Generation based on Information Retrieval and Question Answering.....103**  
*Zhiheng Zhou, Man Lan, Yuanbin Wu and Jun Lang*

<b>Session 5</b>	Word Embedding
<b>Date/Time</b>	Tuesday, 05 December 2017 / 01:30 pm – 03:15 pm
<b>Chair</b>	Prof. <b>Sanath Jayasena</b> , <i>University of Moratuwa, Sri Lanka</i>

- 020 Analysis of Japanese WSD with HiraganaKanji conversion and Context Word Embeddings.....107**  
*Yuki Gumizawa and Kazuhide Yamamoto*
- 021 CBOS: Continuous Bag of Sentences for Learning Sentence Embeddings.....111**  
*Ye Yuan and Yue Zhang*
- 031 Analyzing word embeddings and improving POS tagger of Tigrinya.....115**  
*Yemane Tedla and Kazuhide Yamamoto*
- 033 Bilingual Word Embedding with Sentence Similarity Constraint for Machine Translation.....119**  
*Kui Wu, Xuancong Wang and AiTi Aw*
- 054 Domain Adaption Based on LDA and Word Embedding in SMT.....123**  
*Shaolin Zhu, Yating Yang, Xiao Li, Tonghai Jiang, Lei Wang, Xi Zhou and Chenggang Mi*
- 143 Constructing an Enriched Domain Taxonomy for Hindi using Word Embeddings.....127**  
*Vaishakh K, Pravalika A, Sowmya Kamath S and Geetha V*
- 006 Neural Domain Adaptation for Chinese Word Segmentation.....131**  
*Zuyi Bao, Si Li, Weiran Xu and Sheng Gao*

<b>Session 6</b>	Linguistics
<b>Date/Time</b>	Tuesday, 05 December 2017 / 01:30 pm – 03:15 pm
<b>Chair</b>	Prof. <b>Shili Ge</b> , <i>Guangdong University of Foreign Studies, China</i>

- 066 A Quantitative Study of the Weakening of Emotional Words' Emotional Intensity in "X-yixia" Construction in the Perspective of Construction Coercion.....135**  
*Chenghao Zhu and Pengyuan Liu*
- 007 An Optimality Theory Solution to Disyllabic Tone Sandhi and Neutral Tone in Yichang Dialect.....140**  
*Yan Li*
- 132 Embedding Wikipedia Title Based on Its Wikipedia Text and Categories.....146**  
*Chi-Yen Chen and Wei-Yun Ma*
- 069 A Study on the Reduplication of Chinese Classifiers.....150**  
*Fengcun An and Lei Zhao*
- 129 The Automatic Extraction of common-used Adverbs for Teaching Chinese as Second Language.....155**  
*Zhimin Wang and Meiyu Wang*
- 149 A Multi-dimensional Analysis of Deception.....160**  
*Qi Su*

<b>Session 7</b>	Neural machine translation
<b>Date/Time</b>	Tuesday, 05 December 2017 / 03:45 pm – 05:30 pm
<b>Chair</b>	Dr. Jun Lang, Lazada Group, Singapore

- 068 Recursive Annotations for Attention-Based Neural Machine Translation.....164**  
*Shaolin Ye and Wu Guo*
- 077 Controlling Byte Pair Encoding for Neural Machine Translation.....168**  
*Alfred John Tacorda, Marvin John Ignacio, Nathaniel Oco and Rachel Edita Roxas*
- 088 Improving Character-level Japanese-Chinese Neural Machine Translation with Radicals as an Additional Input Feature.....172**  
*Jinyi Zhang and Tadahiro Matsumoto*
- 089 Exploration of Chinese-Uyghur Neural Machine Translation.....176**  
*Gulnigar Mahmut, Rehmulla Memet, Mewlude Nijat and Askar Hamdulla*
- 096 Language Post Positioned Characteristic Based Chinese-Vietnamese Statistical Machine Translation Method.....180**  
*Jianyalin He, Zhengtao Yu, changtao lv, Hua Lai, Shengxiang Gao and Yang Zhang*
- 114 Error Analysis of Chinese-English Machine Translation on the Clause-Complex Level.....185**  
*Xiaoping Lin, Shili Ge and Rou Song*
- 081 Neural Machine Translation for Sinhala and Tamil Languages.....189**  
*Pasindu Tennage, Prabath Sandaruwan, Malith Thilakarathne, Achini Herath, Surangika Ranathunga, Sanath Jayasena and Gihan Dias*

<b>Session 8</b>	Speech processing
<b>Date/Time</b>	Tuesday, 05 December 2017 / 03:45 pm – 05:30 pm
<b>Chair</b>	Dr. Lei Wang, Institute for Infocomm Research, Singapore

- 032 Study on the Aspirated Characteristics of Chinese Mandarin Consonant.....193**  
*Shiliang Lyu and Luxin Zhou*
- 060 Improving Air Traffic Control Speech Intelligibility by Reducing Speaking Rate Effectively.....197**  
*Nana Hou, Xiaohai Tian, Eng Siong Chng, Bin Ma and Haizhou Li*
- 153 A Light-weight Method of Building an LSTM-RNN-based Bilingual TTS System.....201**  
*Huaiping Ming, Yanfeng Lu, Zhengchen Zhang and Minghui Dong*
- 109 Word Polarity Detection using Syllable Features for Manipuri Language.....206**  
*Loitongbam Gyanendro Singh and Sanasam Ranbir Singh*
- 117 A Review of the Mandarin-English Code-switching Corpus: SEAME.....210**  
*Grandee Lee, Thi-Nga Ho, Eng Siong Chng and Haizhou Li*
- 139 Effect of Language Independent Transcribers on Spoken Language Identification for Different Indian Languages.....214**  
*Rajlakshmi Saikia, Sanasam Ranbir Singh and Priyankoo Sarmah*
- 142 Adapting monolingual resources for code-mixed Hindi-English speech recognition.....218**  
*Ayushi Pandey, Brij Mohan Lal Srivastava and Suryakanth V Gangashetty*

<b>Session 9</b>	Under-resourced language studies
<b>Date/Time</b>	Tuesday, 05 December 2017 / 03:45 pm – 05:30 pm
<b>Chair</b>	Prof. <b>Yan Li</b> , <i>Shaanxi Normal University, China</i>

- 030 Variational Grid Setting Network.....222**  
*Yu-Neng Chuang, Zi-yu Huang and Yen-Lung Tsai*
- 035 Compiling a Text Re-Use Detection Corpus from Scientific Papers with Semi-Real Cases of Plagiarism.....227**  
*salar mohtaj, Habibollah Asghari and Vahid Zarrabi*
- 045 A Rule and Statistical Modeling based Stem Extraction Method for Kazakh Words.....231**  
*Rehmutulla Memet, Gulnigar Mahmut, Mewlude Nijat and Askar Hamdulla*
- 051 HiEnCor: on Mining of a Hi-En General Purpose Parallel Corpus from the Web.....235**  
*Arjun Das, Utpal Garain, Ravindra Kumar and Apurbalal Senapati*
- 086 Chinese-Thai Cross-Language Topic Extraction and Alignment.....239**  
*Xia Li, Zihang Zeng, Jianshu Zhang and Shengyi Jiang*
- 128 “Nee Intention enti?” Towards Dialog Act Recognition in Code-Mixed Conversations.....243**  
*Divya Sai Jitta, Khyathi Raghavi Chandu, Harsha Pamidipalli and Radhika Mamidi*

<b>Session 10</b>	Text mining I
<b>Date/Time</b>	Wednesday, 06 December 2017 / 10:30 am – 12:00 pm
<b>Chair</b>	Prof. <b>Xia Li</b> , <i>Guangdong University of Foreign Studies, China</i>

- 044 Understanding Explicit Arithmetic Word Problems and Explicit Plane Geometry Problems Using Syntax-Semantics Models.....247**  
*Xinguo Yu, Wenbin Gan and Mingshu Wang*
- 055 Simple and sophisticated inling summary generation based on encoder-decoder model and transfer learning.....252**  
*Yuuki Tagawa and Kazutaka Shimada*
- 084 A Joint Framework for Entity Discovery and Linking in Chinese Questions.....256**  
*Ziqi Lin, Wancheng Ni, Haidong Zhang, Yu Liu and Yiping Yang*
- 101 Multi-Document News Summarization via Paragraph Embedding and Density Peak Clustering.....260**  
*Baoyan Wang, Jian Zhang, Fanggui Ding and Yuexian Zou*
- 120 Semantic-Frame Representation for Event Detection on Twitter.....264**  
*Yanxia Qin, Yue Zhang, Min Zhang and Dequan zheng*
- 131 Exploring Semantic Content to User Profiling for User Cluster-based Collaborative Point-of-Interest Recommender System.....268**  
*Yuhuan Xiu, Man Lan, Yuanbin Wu and Jun Lang*

<b>Session 11</b>	Machine learning for natural language II
<b>Date/Time</b>	Wednesday, 06 December 2017 / 10:30 am – 12:00 pm
<b>Chair</b>	Prof. <b>Kazuhide Yamamoto</b> , <i>Nagaoka University of Technology, Japan</i>

- 034 On the Use of Machine Translation-Based Approaches for Vietnamese Diacritic Restoration.....272**  
*Thai-Hoang Pham, Xuan-Khoai Pham and Phuong Le-Hong*
- 078 Filipino and English Clickbait Detection Using a Long Short Term Memory Recurrent Neural Network.....276**  
*Philogene Kyle Dimpas, Royce Vincent Po and Mary Jane Sabellano*
- 106 Transition-based Dependency Parser with Postponed Determinations for Japanese Sentences.....281**  
*Xiaobo Xi and Akihiro Inokuchi*
- 127 Chinese Hedge Scope Detection Based on Phrase Semantic Representation.....285**  
*Huiwei Zhou, Shixian Ning, Yunlong Yang, Zhuang Liu and Junli Xu*
- 150 The Singular Value Decomposition-based Anchor Word Selection Method for Separable Nonnegative Matrix Factorization.....289**  
*Delano Novrilianto, Hendri Murfi and Ari Wibowo*
- 103 Qualitative data analysis of disaster risk reduction suggestions assisted by topic modeling and word2vec.....293**  
*Ken Gorro, Jeffrey Rosario Ancheta, Nathaniel Oco, Rachel Edita Roxas, Mary Jane Sabellano, Brandie Nonnecke, Shrestha Mohanty, Camille Crittenden and Ken Goldberg*

<b>Session 12</b>	Text mining II
<b>Date/Time</b>	Wednesday, 06 December 2017 / 01:30 pm – 03:15 pm
<b>Chair</b>	Dr. <b>Kui Wu</b> , <i>Institute for Infocomm Research, Singapore</i>

- 040 Extraction of Indonesian and English Parallel Sentences from Movie Subtitles.....298**  
*Boon Hong Yeo, Ai Ti Aw and Xuancong Wang*
- 072 Supervised Learning for Robust Term Extraction.....302**  
*Yu Yuan, Jie Gao and Yue Zhang*
- 091 Non-Entity Event Argument Extraction on Structural Representation.....306**  
*Yiting Liu and Peifeng Li*
- 010 A Simple Yet Effective Method for Summarizing Microblogging Users with their Representative Tweets.....310**  
*Shuangyong Song, Yao Meng, Zhiwei Shi, Zhongguang Zheng and Haiqing Chen*
- 105 Using Topic Analysis Techniques to Support Comprehensive Research Paper Searches.....314**  
*Satoshi Fukuda and Yoichi Tomiura*
- 138 Extractive Text Summarisation in Hindi.....318**  
*Sakshee Vijay, Vartika Rai, Sorabh Gupta, Anshuman Vijayvargia and Dipti Misra Sharma*
- 152 Mining Features for Web NER Model Construction based on Distant Learning.....322**  
*Chien-Lung Chou and Chia-Hui Chang*



<b>Session 13</b>	Word sense disambiguation
<b>Date/Time</b>	Wednesday, 06 December 2017 / 01:30 pm – 03:15 pm
<b>Chair</b>	Dr. <b>Ridong Jiang</b> , <i>Institute for Infocomm Research, Singapore</i>

- 015 Co-Occurrence Semantic Knowledge Base Construction for Abbreviation Disambiguation.....326**  
*Shuangyong Song, Qingliang Miao, Zhiwei Shi, Yao Meng and Haiqing Chen*
- 111 Japanese-Chinese Machine Translation for the Japanese Case Particle "de".....330**  
*Jinyi Zhang and Tadahiro Matsumoto*
- 027 Implicit Discourse Relation Identification based on Tree Structure Neural Network.....334**  
*Ruiying Geng, Ping Jian, Yingxue Zhang and Heyan Huang*
- 067 Joint Bi-Affine Parsing and Semantic Role Labeling.....338**  
*Peng Shi and Yue Zhang*
- 070 Word Sense Disambiguation of Adjectives using Dependency Structure and Degree of Association Between Sentences.....342**  
*Kenichi Mishina, Seiji Tsuchiya and Hirokazu Watabe*
- 110 Analysis of Literal and Metaphorical Senses Based on Diachronic Word Embeddings.....346**  
*Yuxiang Jia, Yi Zheng, Hongying Zan and Zhimin Wang*
- 135 Improving Word and Sense Embedding with Hierarchical Semantic Relations.....350**  
*Yow-Ting Shiue and Wei-Yun Ma*

<b>Session 14</b>	NLP applications
<b>Date/Time</b>	Wednesday, 06 December 2017 / 03:45 pm – 05:15 pm
<b>Chair</b>	Dr. <b>Zhengchen Zhang</b> , <i>Institute for Infocomm Research, Singapore</i>

- 037 The Chinese Vietnamese Bilingual News Event Ranking Method Based on Attribute Association Graph.....354**  
*Mingwei Zhu, Zhengtao YU, Guangshun Qin, Hua Lai and Shengxiang Gao*
- 042 Complement of Incomplete Task Results for Real-time Crowdsourcing Interpretation.....359**  
*Takeaki Shionome, Hirotaka Hashimoto, Jianwei Zhang, Yuhki Shiraishi, Daisuke Wakatsuki, Yohei Seki and Atsuyuki Morishima*
- 050 Sentence simplification with core vocabulary.....363**  
*Takumi Maruyama and Kazuhide Yamamoto*
- 075 Neural Architecture for Tibetan Word Segmentation.....367**  
*Mengzhu Chen, Shengjie Zhao and Kai Yang*
- 095 Hybrid Answer Selection Model for Non-Factoid Question Answering.....371**  
*Rongqiang Ma, Jian Zhang, Miao Li, Lei Chen and Jin Gao*
- 112 Named Entity Transliteration with Sequence-to-Sequence Neural Network.....374**  
*Zhongwei Li, Eng Siong Chng and Haizhou Li*

<b>Session 15</b>	Sentiment Analysis
<b>Date/Time</b>	Wednesday, 06 December 2017 / 03:45 pm – 05:15 pm
<b>Chair</b>	Dr. <b>Xuancong Wang</b> , <i>Institute for Infocomm Research, Singapore</i>

- 028 A Hierarchical LSTM Model with Multiple Features for Sentiment Analysis of Sina Weibo Texts.....379**  
*Shumin Shi, Meng Zhao, Jun Guan, Yaxuan Li and Heyan Huang*
- 049 Unsupervised Aspect-Based Sentiment Analysis on Indonesian Restaurant Reviews.....383**  
*Dhanang Hadhi Sasmita, Alfian F. Wicaksono, Samuel Louvan and Mirna Adriani*
- 056 Fine-grained Sentiment Analysis with 32 Dimensions.....387**  
*Xianchao Wu, Hang Tong and Momo Klyen*
- 122 InSet Lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs.....391**  
*Fajri Koto and Gemala Y. Rahmaningtyas*
- 126 A Microblog Dataset for Tibetan Sentiment Analysis.....395**  
*Yong Cuo, Xiaodong Shi, Nyima Trashi and Yidong Chen*
- 146 Towards Question Identification from Online Healthcare Consultation Forum Post in Bahasa.....399**  
*Rahmad Mahendra, Abid Nurul Hakim and Mirna Adriani*