

# **2022 International Conference on Asian Language Processing (IALP 2022)**

**Singapore  
27-28 October 2022**



**IEEE Catalog Number: CFP2244I-POD  
ISBN: 978-1-6654-7675-1**

**Copyright © 2022 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:	CFP2244I-POD
ISBN (Print-On-Demand):	978-1-6654-7675-1
ISBN (Online):	978-1-6654-7674-4

**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

## Table of Content

<b>About the Conference .....</b>	<b>X</b>
<b>Message from the Conference Chair .....</b>	<b>XI</b>
<b>Message from the Program Chairs .....</b>	<b>XII</b>
<b>Conference Committees .....</b>	<b>XIII</b>
<b>Keynote Speech.....</b>	<b>XV</b>
<b>Author Index .....</b>	<b>XVI</b>
<b>Papers</b>	
<b>1: Semantics and Sentiment Analysis</b>	
<b>S1.1 Distant Supervision For Chinese Semantic Role Labeling</b> <i>Yi Zhang, Guirong Wang and Endong Xun .....</i>	<b>1</b>
<b>S1.2 A Self-Reliant finite Automata for Reduplication Detection</b> <i>Apurbalal Senapati.....</i>	<b>9</b>
<b>S1.3 Theft or Felony? Task Augmentation for Criminal Amount Calculation in Judgment Documents</b> <i>Fei Li, Yanxia Qin and Shoujian Yu .....</i>	<b>14</b>
<b>S1.4 Research on Multi-feature Expert Disambiguation of Same Name fused Personal Experience</b> <i>Shangmei Li, Yangsen Zhang, Xiang Chen, Han Chen and Gaijuan Huang .....</i>	<b>20</b>
<b>S1.5 PosWSD: Low-Resource Word Sense Disambiguation Model using Part Of Speech Information</b> <i>Yazhen Chen, Jian Zhang and Qipeng He .....</i>	<b>26</b>
<b>S1.6 The Construction of the Semantic Collocation Database of Verb-Complement Structure in Modern Chinese based on a Large-scale Chinese Chunkbank</b> <i>Tian Shao, ShiQuan Zhai, Gaoqi Rao and Endong Xun .....</i>	<b>32</b>
<b>S1.7 Improving Sentiment Analysis By Emotion Lexicon Approach on Vietnamese Texts</b> <i>An Long Doan and Son T. Luu .....</i>	<b>39</b>
<b>S1.8 Distillation Language Adversarial Network for Cross-lingual Sentiment Analysis</b> <i>Deheng Wang, Aimin Yang, Yongmei Zhou, Fenfang Xie, Zhouhao Ouyang and Sancheng Peng .....</i>	<b>45</b>
<b>S1.9 Emoji Prediction from Arabic Sentence</b> <i>Takua Mokhamed and Ashraf Elnagar .....</i>	<b>51</b>
<b>S1.10 Is Twitter an Echo Chamber? Connecting Online Public Sentiments to Actual Results From the 2019 Philippine Midterm Elections</b> <i>Lamar Clarence Cruz, Jessica Nicole Dela Cruz, Shane Francis Maglangit, Mico Magtira, Joseph Marvin Imperial and Ramon Rodriguez .....</i>	<b>57</b>

<b>S1.11</b> Aspect-based Sentiment Analysis with Dependency Relation Graph Convolutional Network <i>Yadong Wang, Chen Liu, Jinge Xie, Songhua Yang, Yuxiang Jia and Hongying Zan</i> .....	63
--	----

**2: Phonetics, Morphology and Phonology**

<b>S2.1</b> On Physical Thoughts of Language Change <i>Xiaochuan Song</i> .....	69
--	----

<b>S2.2</b> Counter-factivity Investigation of caiguai <i>Baoshuang Mao and Zhimin Wang</i> .....	75
--	----

<b>S2.3</b> The Belongings of Zhushan Dialect in Hubei Province <i>Yan Li</i> .....	81
--	----

<b>S2.4</b> Literary versus Colloquial Distinction and Phonetic Strata in Huazhou Dialect <i>Yingping Wang</i> .....	87
---	----

<b>S2.5</b> A Study of the Language Contact between Wuqi Dialect and Putonghua <i>Yanmei Liu</i> .....	93
---	----

<b>S2.6</b> The Production of English Question Intonation by Cantonese and Mandarin Speakers <i>Qi Xie</i> .....	98
---	----

<b>S2.7</b> Pairwise T-Test: Identification of Tonal Relations <i>Yishan Huang</i> .....	104
---	-----

<b>S2.8</b> Construction and Analysis of Mongolian Word Co-occurrence Networks <i>Lingxiong Bao and Dahubaiyila</i> .....	110
--	-----

<b>S2.9</b> Identifying Relation Between Miriek and Kenyah Badeng Language by Using Morphological Analyzer <i>Muhammad Abdullah Yusof, Suhaila Saeed and Sarah Samson Juan</i> .....	116
---	-----

<b>S2.10</b> Boundary Feature Analysis and Automatic Recognition of Indirect Speech in the Korean News Field <i>Huijie Zhang, Wanjun Zeng and Yude Bi</i> .....	122
--	-----

**3: Character Recognition and Speech Processing**

<b>S3.1</b> Research on the processing of verb-noun anomalous collocation with components ellipsis <i>Mengxiang Wang</i> .....	128
---	-----

<b>S3.2</b> Handwriting recognition in Tibetan based on active learning strategy <i>Lili Wang, Sen Bao, Yanshan He and Hongwu Yang</i> .....	132
---	-----

<b>S3.3</b> Isolated Kannada Character Recognition using Densely Connected Convolutional Network <i>Sandhya S and Geetha V</i> .....	137
---	-----

<b>S3.4</b> Adapting the Tesseract Open-Source OCR Engine for Tamil and Sinhala Legacy Fonts and Creating a Parallel Corpus for Tamil-Sinhala-English <i>Charangan Vasantharajan, Laksika Tharmalingam and Uthayasanker Thayasivam</i> .....	143
---	-----

<b>S3.5</b> Research on Syllable-Based Language Model in Malay Speech Recognition <i>Xiangfeng Wei, Quan Zhang and Yi Yuan</i> .....	150
<b>S3.6</b> End-to-End Myanmar Speech Recognition with Human-Machine Cooperation <i>Faliang Wang, Yiling Yang and Jian Yang</i> .....	156
<b>S3.7</b> MFCC-based deep convolutional neural network for audio depression recognition <i>Yafan Wang, Xiaoyong Lu and Daimin Shi</i> .....	162
<b>S3.8</b> Meeting Record Generation Based on Hierarchical Decomposition Encoding <i>Qingqing Chen, Jian Zhang and Mingzi Fu</i> .....	167
<b>S3.9</b> An Investigation of Magnitude-Based and Phase-Based Features for Large-Scale Speaker Identification <i>Mengjie Du, Zhijunyi Yang, Rongfeng Su, Nan Yan and Lan Wang</i> .....	173
<b>S3.10</b> Unified Voice Embedding through Multi-task Learning <i>Jenarthan Rajenthiran, Lakshikka Sithamparanathan, Saranya Uthayakumar and Uthayasanker Thayasivam</i> .....	178
<b>S3.11</b> MnTTS: An Open-Source Mongolian Text-to-Speech Synthesis Dataset and Accompanied Baseline <i>Yifan Hu, Pengkai Yin, Rui Liu, Feilong Bao and Guanglai Gao</i> .....	184
<b>4: Text Classification, Information Extraction and Named Entity Recognition</b>	
<b>S4.1</b> A New Method for Predicting Severity Level of Dysarthric Speech Based on Joint Feature-Sample Selection using Audio-Visual Data <i>Shangjun Lu, Xiaoxia Du, Juan Liu, Yu-Mei Zhang, Shaofeng Zhao, Rongfeng Su, Lan Wang and Nan Yan</i> .....	190
<b>S4.2</b> Transfer Learning Across Arabic Dialects for Offensive Language Detection <i>Fatemah Husain and Ozlem Uzuner</i> .....	196
<b>S4.3</b> A Multi-label Classification Algorithm for Non-standard Text <i>Shihong Chen, Qichao Kuang, Xiaohui Yu, Suifeng Li and Ruoyao Ding</i> .....	206
<b>S4.4</b> Multi-task Learning Model for Detecting Internet Slang Words with Two-Layer Annotation <i>Yohei Seki and Yihong Liu</i> .....	212
<b>S4.5</b> Text Classification of Modern Mongolian Documents using BERT models <i>Garmaabazar Khaltarkhuu, Biligsaikhan Batjargal and Akira Maeda</i> .....	219
<b>S4.6</b> Petroleum Exploration and Development Text Triplet Extraction Based on Deep Learning <i>Jianli Li, Hankiz Yilahun and Askar Hamdulla</i> .....	225
<b>S4.7</b> Text-Level Contrastive Learning for Scene Text Recognition <i>Junbin Zhuang, Yixuan Ren, Xia Li and Zhanpeng Liang</i> .....	231

<b>S4.8</b> Chinese Futures Market Information Extraction and Classification Model Combined with HMM and SVM <i>Zhiming Yang, Heguo Yang, Guoping Zhao and Lingling Mu</i> .....	237
<b>S4.9</b> Named Entity Recognition of Protein Complex Incorporating Syntactic Information <i>Junsong Ruan, Hailin Zhang, Hai Zhu and Ruoyao Ding</i> .....	241
<b>S4.10</b> Biomedical Named Entity Recognition Based on MCBERT <i>Sai Wang, Hankiz Yilahun and Askar Hamdulla</i> .....	247
<b>S4.11</b> Biomedical Entity Linking Based on Global and Local Feature Fusion <i>Turdi Tohti, Mamatjan Abdurxit and Askar Hamdulla</i> .....	253
<b>S4.12</b> Pretrained Models with Adversarial Training for Named Entity Recognition in Scientific Text <i>Hangchao Ma, You Zhang and Jin Wang</i> .....	259
<b>5: Dialogue, Q&amp;A and Text Generation</b>	
<b>S5.1</b> Domain Specific Intent Classification with BiLSTM <i>Jayani Hellarawa and Uthayasanker Thayasivam</i> .....	265
<b>S5.2</b> Joint Training Model of Intent Detection and Slot Filling for Multi Granularity Implicit Guidance <i>Bin Li, Weihua Wang and Feilong Bao</i> .....	271
<b>S5.3</b> Constraint-based Paraphrase Generation Model and it's Application on Message Extraction of Commodity Futures News <i>Yintao Jia, Xiaodong Ning, Liqin Gong and Lingling Mu</i> .....	275
<b>S5.4</b> Paragraph-level Tibetan Question Generation for Machine Reading Comprehension <i>Sisi Liu, Chaofan Chen and Yuan Sun</i> .....	280
<b>S5.5</b> Applying Emotional Keyphrase Correlation for Diversity Enhancement in Empathetic Dialogue Response Generation <i>Jeremy Chang and Chung-Hsien Wu</i> .....	286
<b>S5.6</b> Generating Questions for Abstractive Answers with a Multi-Channel Fusion Encoder-Decoder Model <i>Jiazuo Qiu and Deyi Xiong</i> .....	292
<b>S5.7</b> A Hybrid Response Generation Model for an Empathetic Conversational Agent <i>Jacklynn L. Beredo and Ethel C. Ong</i> .....	300
<b>S5.8</b> Using Synthetic Data to Train a Conversational Response Generation Model in Low Resource Settings <i>Denzel Adrian Co, Schuyler Ng, Gabriel Louis Tan, Adrian Paule Ty, Jan Blaise Cruz and Charibeth Cheng</i> .....	306
<b>S5.9</b> A Survey of Machine Reading Comprehension Methods <i>Xiaobo Xu, Turdi Tohti and Askar Hamdulla</i> .....	312

<b>S5.10</b> Answer Selection Method Based on BERT and Parallel Multi-Channel Convolution <i>Jianlong Li, Yangsen Zhang and Jiang Miao</i> .....	318
<b>S5.11</b> Domain specific Question & Answer generation in Tamil <i>Rubika Murugathas and Uthayasanker Thayasivam</i> .....	323
<b>6: Language Education</b>	
<b>S6.1</b> The Effect of Chinese Perfective Aspect Marker <i>le</i> on the Simple Past Use in English Interlanguage <i>Peicui Zhang, Jie Li and Huibin Zhuang</i> .....	329
<b>S6.2</b> The Types and Strategies of Chinese Learners' Vocabulary Errors Based on the Free Output <i>Bingbing Yang and Zhimin Wang</i> .....	335
<b>S6.3</b> Oral Materials in the Legal Workplace Applied in the Automatic Push of Chinese Teaching Material <i>Bingbing Yang and Zhimin Wang</i> .....	341
<b>S6.4</b> A Survey and Analysis on Chinese Degree Adverbs Acquisition of International Students <i>Liangjie Yuan and Zhimin Wang</i> .....	347
<b>S6.5</b> Research on the Evaluation of the Classical Chinese Difficulty in the Compulsory Education Stage <i>Kun Ma, Zhiying Liu, Lijiao Yang, Ying Sun, Yujun Wang and Ziliang Qiu</i> .....	353
<b>S6.6</b> The Contributions of Initials and Finals to L2 Chinese Comprehensibility Based on Functional Load Principle <i>Rian Bao, Linkai Peng and Jinsong Zhang</i> .....	358
<b>S6.7</b> Effects of Imitation Training Methods on Chinese EFL Learners' Production of English Intonation <i>Zhilin Huang, Huijie Yu, Haoqing Wu and Ying Chen</i> .....	363
<b>S6.8</b> Acoustical Analysis on English Sibilant Fricative /s, ʃ/ Produced by Native Chinese Speakers <i>Yuhong Sun, Jian Gong, Kai Sheng, Weizhong Zhang, William Bellamy and Xiaoli Ji</i> .....	369
<b>S6.9</b> A Formal and Functional Analysis of Repetition in Daily Conversation <i>Yanjiao Li</i> .....	374
<b>S6.10</b> The Importance of Lexical Tone for Sentence Understanding: Utilizing Functional Load Principle to Simulate Comprehension Process <i>Rian Bao, Linkai Peng, Yingming Gao and Jinsong Zhang</i> .....	379
<b>S6.11</b> Acoustic Analysis of the Acquisition of Plosives and Affricates by Preschool Teachers <i>Zili Wang and Yuan Jia</i> .....	384
<b>S6.12</b> Influence of Chinese Er Suffixation on American English R-colored Vowels by Northeast Chinese EFL Learners <i>Jiwei Guo, Hui Feng and Yuan Jia</i> .....	389

**7: Corpus, Knowledge Base and NLP Application**

**S7.1** Software Defect Detection Method Based on Graph Structure and Deep Neural Network  
*Tianrun Qu, Wei Liu, Weining Zheng and Wenxin Tao* ..... 395

**S7.2** Graph-based Model Using Text Simplification for Readability Assessment  
*Rui Xu, Wenjing Pan, Canhua Chen, XiaoYin Chen, Shilin Lin and Xia Li* ..... 401

**S7.3** Tapping the Potential of Coherence and Syntactic Features in Neural Models for Automatic Essay Scoring  
*Xinying Qiu, Shuxuan Liao, Jiajun Xie and Jian-Yun Nie* ..... 407

**S7.4** Enhancing Automatic Readability Assessment with Verb Frame Features  
*Tianyuan CAI, Ho Hung LIM, John S. Y. LEE and Meichun LIU* ..... 413

**S7.5** Modern Chinese Diachronic Corpus Construction and the Language Development Based on a 77-year Corpus  
*Zixi Li, Tianle Gao, Guojing Huang and Gaoqi Rao* ..... 419

**S7.6** Gender Stereotypes in TCSOL Dialogue Corpus  
*Jiali Li, Shucheng Zhu, Ying Liu and Pengyuan Liu* ..... 427

**S7.7** Automatic Generation and Evaluation of Chinese Grammar Proofreading Corpus  
*Mei Zhang, Lijian Pan, Jianyong Duan, Zhitong Xu and Lishan Xu* ..... 433

**S7.8** InDEX: Indonesian Idiom and Expression Dataset for Cloze Test  
*Xinying Qiu and Guofeng Shi* ..... 439

**S7.9** Rule-based Logical Reasoning Knowledge Extraction of Chinese Legislative Text  
*Li Li, Houfeng Wang and Yang Liu* ..... 445

**S7.10** Research on the Construction of Non-subject-object Argument Knowledge Base of Chinese Verbs  
*Chengwen Wang and Endong Xun* ..... 451

**S7.11** Research on Partition Block-based Multi-source Knowledge Fusion for Knowledge Graph Construction  
*Yu Song, Dongming Dai, Chenxin Hu, Mingyu Gui, Hongying Zan and Kunli Zhang* ..... 458

**S7.12** An Adaptive Session Selection Method for Recommendation System  
*Dongyang Liu, Shumin Shi and Sichen Liu* ..... 464

**8: Machine Translation**

**S8.1** How Much Data is Sufficient for Neural Transliteration?  
*Zhijuan Wang, Zhuo Li and Xiaobing Zhao* ..... 470

**S8.2** Sub-word based unsupervised bilingual dictionary induction for Chinese-Uyghur  
*Anwar Aysa, Mijit Ablimit, Hankiz Yilahun and Askar Hamdulla* ..... 476



<b>S8.3</b> NEB-Filter: A Simple but Effective Filter based on Named Entity Boundaries for Low-Resource Cross-Lingual NER <i>Linzhen Jian, Ping Jian and Weilun Fei</i> .....	482
<b>S8.4</b> Building Indonesian Dependency Parser Using Cross-lingual Transfer Learning <i>Andhika Yusup Maulana, Ika Alfina and Kurniawati Azizah</i> .....	488
<b>S8.5</b> Multilingual Neural Machine Transliteration with Adaptive Segmentation Schemes <i>Haoran Sun, Yikun Lei and Deyi Xiong</i> .....	494
<b>S8.6</b> Ensemble Machine Translation to Filter Low Quality Corpus <i>Wuying Liu and Lin Wang</i> .....	500