

2018 International Conference on Asian Language Processing (IALP 2018)

**Bandung, Indonesia
15-17 November 2018**



**IEEE Catalog Number: CFP1844I-POD
ISBN: 978-1-5386-8298-2**

**Copyright © 2018 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: | CFP1844I-POD |
| ISBN (Print-On-Demand): | 978-1-5386-8298-2 |
| ISBN (Online): | 978-1-7281-1175-9 |

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

CURRAN ASSOCIATES INC.
proceedings
.com

TABLE OF CONTENTS

| | |
|--|------------|
| DISYLLABIC TONE SANDHI AND NEUTRAL TONE PATTERNS IN YICHANG DIALECT | 1 |
| <i>Yan Li ; Zhiyi Wu</i> | |
| SYNTACTIC-SEMANTIC KNOWLEDGE REPRESENTATION FRAMEWORK FOR KOREAN VERBS AND ITS WORKING MECHANISMS | 8 |
| <i>Yude Bi ; Jianguo Xiong</i> | |
| LINGUISTIC DIVERGENCE OF SINHALA AND TAMIL LANGUAGES IN MACHINE TRANSLATION | 13 |
| <i>W.S.N. Dilshani ; S. Yashothara ; R. T. Uthayasanker ; S. Jayasena</i> | |
| RUDE-WORDS DETECTION FOR INDONESIAN SPEECH USING SUPPORT VECTOR MACHINE | 19 |
| <i>Sashi Novitasari ; Dessi Puji Lestari ; Sakriani Sakti ; Ayu Purwarianti</i> | |
| LEARNING HOW TO SELF-LEARN: ENHANCING SELF-TRAINING USING NEURAL REINFORCEMENT LEARNING | 25 |
| <i>Chenhua Chen ; Yue Zhang ; Yuze Gao</i> | |
| A HYBRID ALGORITHM FOR TEXT CLASSIFICATION BASED ON CNN-BLSTM WITH ATTENTION | 31 |
| <i>Lei Fu ; Zhaoxia Yin ; Xin Wang ; Yi Liu</i> | |
| INVESTIGATING BI-LSTM AND CRF WITH POS TAG EMBEDDING FOR INDONESIAN NAMED ENTITY TAGGER..... | 35 |
| <i>Devin Hoesen ; Ayu Purwarianti</i> | |
| INDONESIA HATE SPEECH DETECTION USING DEEP LEARNING..... | 39 |
| <i>Taufic Leonardo Sutejo ; Dessi Puji Lestari</i> | |
| RATING PREDICTOR: SENTIMENT ANALYSIS OF PRODUCT REVIEWS IN ARABIC | 44 |
| <i>Fouzi Harrage ; Abdulmalik Als Salman ; Alaa Alqahtani</i> | |
| REVIEW RATING WITH BOTH POSITIVE AND NEGATIVE SENTIMENT INTENSITY MEASUREMENTS | 50 |
| <i>Lu Zhang ; Yijie Zhang ; Shoushan Li ; Guodong Zhou</i> | |
| COMPARATIVE STUDY ON LANGUAGE RULE BASED METHODS FOR ASPECT EXTRACTION IN SENTIMENT ANALYSIS | 56 |
| <i>Fariska Zakhralatwa Ruskanda ; Dwi Hendratmo Widyantoro ; Ayu Purwarianti</i> | |
| ASPECT DETECTION AND SENTIMENT CLASSIFICATION USING DEEP NEURAL NETWORK FOR INDONESIAN ASPECT-BASED SENTIMENT ANALYSIS..... | 62 |
| <i>Arfinda Ilmania ; Abdurrahman ; Samuel Cahyawijaya ; Ayu Purwarianti</i> | |
| INVESTIGATING MANIFOLD LEARNING TECHNIQUE FOR ROBUST SPEECH RECOGNITION..... | 68 |
| <i>Bi-Cheng Yan ; Chin-Hong Shih ; Berlin Chen ; Shih-Hung Liu</i> | |
| INVESTIGATING FOR PUNCTUATION PREDICTION IN CHINESE SPEECH TRANSCRIPTIONS..... | 74 |
| <i>Xin Liu ; Yi Liu ; Xiao Song</i> | |
| SPEECH TO TEXT OF PATIENT COMPLAINTS FOR BAHASA INDONESIA | 79 |
| <i>Teguh Puji Laksono ; Ahmad Fathan Hidayatullah ; Chanifah Indah Ratnasari</i> | |
| RESEARCH ON TRANSFER LEARNING FOR KHALKHA MONGOLIAN SPEECH RECOGNITION BASED ON TDNN..... | 85 |
| <i>Linyan Shi ; Feilong Bao ; Yonghe Wang ; Guanglai Gao</i> | |
| EMOTION CLASSIFICATION ON INDONESIAN TWITTER DATASET | 90 |
| <i>Mei Silviana Saputri ; Rahmad Mahendra ; Mirna Adriani</i> | |
| STANCE CLASSIFICATION TOWARDS POLITICAL FIGURES ON BLOG WRITING | 96 |
| <i>Rini Jannati ; Rahmad Mahendra ; Cakra Wishnu Wardhana ; Mirna Adriani</i> | |
| CHINESE DIALOGUE ANALYSIS USING MULTI-TASK LEARNING FRAMEWORK | 102 |
| <i>Xuejing Zhang ; Xueqiang Lv ; Qiang Zhou</i> | |
| THE LAW OF SEMANTIC CHANGE OF OPPOSITE COMPOUNDS..... | 108 |
| <i>Yingdi Jiang ; Yonglian Yu ; Xiaoyu Wang ; Zhiying Liu</i> | |
| WARNING AND SUGGESTION SYSTEM ON SYNTAX TREE MAKER APPLICATION | 113 |
| <i>Mukhtar Haris ; Moch Arif Bijaksana ; Totok Suhardijanto</i> | |
| A STUDY OF THE COLOR CATEGORY AND COLLOCATION EVOLUTION OF CLASSICAL POETRY | 118 |
| <i>Yonglian Yu ; Yingdi Jiang</i> | |

| | |
|--|-----|
| A STUDY OF CODE-SWITCHING IN CHINESE WEB NOVELS | 123 |
| <i>Qi Su</i> | |
| A HYBRID DEEP LEARNING ARCHITECTURE FOR SENTENCE UNIT DETECTION | 129 |
| <i>Duy-Cat Can ; Thi-Nga Ho ; Eng-Siong Chng</i> | |
| BUILDING MEDISCO: INDONESIAN SPEECH CORPUS FOR MEDICAL DOMAIN | 133 |
| <i>Muhammad Reza Qorib ; Mirna Adriani</i> | |
| AN INVESTIGATION OF WORD EMBEDDINGS WITH DEEP BIDIRECTIONAL LSTM FOR SENTENCE UNIT DETECTION IN AUTOMATIC SPEECH TRANSCRIPTION | 139 |
| <i>Thi-Nga Ho ; Duy-Cat Can ; Engsiong Chng</i> | |
| UTILIZING INDONESIAN ALLOPHONES AND INTRAWORD SHORT PAUSES HANDLING TO IMPROVE PERFORMANCE OF INDONESIAN TEXT-TO-SPEECH | 143 |
| <i>Mohammad Teduh Uliniansyah ; Elvira Nurfadhilah ; Harnum Annisa ; Made Gunawan ; Lyla Ruslana Aini ; Agung Santosa ; Asril Jarin ; Gunarso ; Fara Ayuningtyas ; Hammam Riza</i> | |
| IMPROVING NEURAL MACHINE TRANSLATION WITH NEURAL SENTENCE REWRITING | 147 |
| <i>Tian Wu ; Zhongjun He ; Enhong Chen ; Haifeng Wang</i> | |
| ACTIVE LEARNING FOR NEURAL MACHINE TRANSLATION | 153 |
| <i>Pei Zhang ; Xueying Xu ; Deyi Xiong</i> | |
| TWO EFFECTIVE APPROACHES TO DATA REDUCTION FOR NEURAL MACHINE TRANSLATION: STATIC AND DYNAMIC SENTENCE SELECTION | 159 |
| <i>Xueying Xu ; Shaohui Kuang ; Deyi Xiong</i> | |
| ENHANCING THE QUALITY OF PHRASE-TABLE IN STATISTICAL MACHINE TRANSLATION FOR LESS-COMMON AND LOW-RESOURCE LANGUAGES | 165 |
| <i>Minh-Thuan Nguyen ; Van Tanbui ; Huy-Hien Vu ; Phuong-Thai Nguyen ; Chi-Mai Luong</i> | |
| WHERE IS THE HEAD POSITIONED IN INDONESIAN LANGUAGE?: A CORPUS STUDY OF HEAD DIRECTIONALITY FROM A DEPENDENCY PERSPECTIVE | 171 |
| <i>Lalitia Ansari ; Totok Suhardijanto</i> | |
| EXPLORING LEXICAL DIFFERENCES BETWEEN INDONESIAN AND MALAY | 178 |
| <i>Nankai Lin ; Sihui Fu ; Shengyi Jiang ; Gangqin Zhu ; Yanni Hou</i> | |
| AN ANALYSIS OF JAPANESE NAMED ENTITY RECOGNIZER SPECIALIZED FOR PERSON AND ORGANIZATION ENTITIES | 184 |
| <i>Takashi Inui ; Yuki Nakano</i> | |
| AUTOMATIC RECOGNITION OF TUNE NAMES OF SONG CI-POETRY | 189 |
| <i>Bihua Wang ; Jianyu Zheng ; Yueming Du ; Lijiao Yang</i> | |
| INDONESIAN CORPUS CONSTRUCTING AND TEXT PROCESSING FOR SPEECH SYNTHESIS | 193 |
| <i>Xuan Kong ; Jian Yang</i> | |
| DOMAIN SPECIFIC INTENT CLASSIFICATION OF SINHALA SPEECH DATA | 197 |
| <i>Darshana Buddhika ; Ranula Liyadipita ; Sudeepa Nadeeshan ; Hasini Witharana ; Sanath Javaseena ; Uthayasanker Thayasivam</i> | |
| HYBRID HMM-BLSTM-BASED ACOUSTIC MODELING FOR AUTOMATIC SPEECH RECOGNITION ON QURAN RECITATION | 203 |
| <i>Faza Thirafi ; Dessi Puji Lestari</i> | |
| A DNN-BASED FRAMEWORK FOR CONVERTING SIGN LANGUAGE TO MANDARIN-TIBETAN CROSS-LINGUAL EMOTIONAL SPEECH | 209 |
| <i>Nan Song ; Hongwu Yang ; Tingting Zhang</i> | |
| INDOSUM: A NEW BENCHMARK DATASET FOR INDONESIAN TEXT SUMMARIZATION | 215 |
| <i>Kemal Kurniawan ; Samuel Louvan</i> | |
| LANGUAGE RESOURCE EXTENSION FOR INDONESIAN-CHINESE MACHINE TRANSLATION | 221 |
| <i>Wuying Liu ; Lixian Xiao ; Shengyi Jiang ; Lin Wang</i> | |
| COLLOQUIAL INDONESIAN LEXICON | 226 |
| <i>Nikmatun Aliyah Salsabila ; Yosef Ardhitto Winatmoko ; Ali Akbar Septiandri ; Ade Jamal</i> | |
| BUILDING CANDIDATE MONOLINGUAL PARALLEL CORPUS FROM SCIENTIFIC PAPERS | 230 |
| <i>Ridwan Ilyas ; Dwi H. Widiyantoro ; Masayu Leylia Khodra</i> | |
| LEARNING INDONESIAN FREQUENTLY USED VOCABULARY FROM LARGE-SCALE NEWS | 234 |
| <i>Nankai Lin ; Sihui Fu ; Shengyi Jiang ; Chen Chen ; Lixian Xiao ; Gangqin Zhu</i> | |
| WORD SEGMENTATION FOR JAVANESE CHARACTER USING DICTIONARY, SVM, AND CRF | 240 |
| <i>Dipta Tanaya ; Mirna Adriani</i> | |
| A JAVANESE SYLLABIFIER BASED ON ITS ORTHOGRAPHIC SYSTEM | 244 |
| <i>Lucia D. Krisnawati ; Aditya W. Mahastama</i> | |

| | |
|---|-----|
| USE OF WORD AND CHARACTER N-GRAMS FOR LOW-RESOURCED LOCAL LANGUAGES | 250 |
| <i>Ralph Vincent Regalado ; Abien Fred Agarap ; Renz Iver Baliber ; Arian Yambao ; Charibeth Cheng</i> | |
| TRANSLITERATION OF ENGLISH LOANWORDS AND NAMED-ENTITIES TO MANIPURI: PHONEME VS GRAPHEME REPRESENTATION | 255 |
| <i>Lenin Laitonjam ; Loitongbam Gyanendro Singh ; Sanasam Ranbir Singh</i> | |
| WORD LEVEL LANGUAGE IDENTIFICATION IN ASSAMESE-BENGALI-HINDI-ENGLISH CODE-MIXED SOCIAL MEDIA TEXT | 261 |
| <i>Neelakshi Sarma ; Sanasam Ranbir Singh ; Diganta Goswami</i> | |
| CLASSIFYING TEMPORAL RELATIONS BETWEEN EVENTS BY DEEP BILSTM | 267 |
| <i>Yijie Zhang ; Peifeng Li ; Guodong Zhou</i> | |
| AN OVERVIEW OF NAMED ENTITY RECOGNITION | 273 |
| <i>Peng Sun ; Xuezheng Yang ; Xiaobing Zhao ; Zhijuan Wang</i> | |
| EXTRACTIVE SUMMARIZATION OF DOCUMENTS BY COMBINING SEMANTIC CONTENT AND NON-STRUCTURED FEATURES | 279 |
| <i>Shan Yang ; Yating Yang ; Chenggang Mi ; Yirong Pan ; Lei Wang ; Bo Ma</i> | |
| HYPERNYM-HYPONYM RELATION EXTRACTION FROM INDONESIAN WIKIPEDIA TEXT | 285 |
| <i>Made Nindyatama Nityasya ; Rahmad Mahendra ; Mirna Adriani</i> | |
| RELATION DETECTION FOR INDONESIAN LANGUAGE USING DEEP NEURAL NETWORK - SUPPORT VECTOR MACHINE | 290 |
| <i>Ramos Janoah Hasudungan ; Ayu Purwarianti</i> | |
| TIBETAN WORD SEGMENTATION METHOD BASED ON BILSTM_ CRF MODEL | 297 |
| <i>Lili Wang ; Hongwu Yang</i> | |
| TOWARD A STANDARDIZED AND MORE ACCURATE INDONESIAN PART-OF-SPEECH TAGGING | 303 |
| <i>Kemal Kurniawan ; Alham Fikri Aji</i> | |
| SENTENCE SIMILARITY COMPUTATION BY INTEGRATING SHALLOW AND DEEP INFORMATION | 308 |
| <i>Pengyuan Liu ; Zhijun Zheng ; Qi Su</i> | |
| ATTENTIVE SIAMESE LSTM NETWORK FOR SEMANTIC TEXTUAL SIMILARITY MEASURE | 312 |
| <i>Wei Bao ; Wugedele Bao ; Jinhua Du ; Yuanyuan Yang ; Xiaobing Zhao</i> | |
| THE SEMANTIC ANNOTATION OF THE QURAN CORPUS BASED ON HIERARCHICAL NETWORK OF CONCEPTS THEORY | 318 |
| <i>Zhiying Liu ; Lijiao Yang ; Eric Atwell</i> | |
| IMPROVING PHRASE-BASED STATISTICAL MACHINE TRANSLATION WITH PREPROCESSING TECHNIQUES | 322 |
| <i>S. Yashothara ; R. T. Uthayasanker ; S. Jayasena</i> | |
| NEURAL MACHINE TRANSLATION FOR CEBUANO TO TAGALOG WITH SUBWORD UNIT TRANSLATION | 328 |
| <i>Kristine Mae M. Adlaon ; Nelson Marcos</i> | |
| BUILDING THE INDONESIAN NE DATASET USING WIKIPEDIA AND DBPEDIA WITH ENTITIES EXPANSION METHOD ON DBPEDIA | 334 |
| <i>Haji Dito Murya Alfarohmi ; Moch. Arif Bijaksana</i> | |
| BURMESE WORD SEGMENTATION METHOD AND IMPLEMENTATION BASED ON CRF | 340 |
| <i>Chang'e Ma ; Jian Yang</i> | |
| MULTI-LABEL TOPIC CLASSIFICATION OF HADITH OF BUKHARI (INDONESIAN LANGUAGE TRANSLATION)USING INFORMATION GAIN AND BACKPROPAGATION NEURAL NETWORK | 344 |
| <i>Muhammad Yuslan Abu Bakar ; Adiwijaya ; Said Al Faraby</i> | |
| AUTHOR-TOPIC MODELLING FOR REVIEWER ASSIGNMENT OF SCIENTIFIC PAPERS IN BAHASA INDONESIA | 351 |
| <i>Renny Pradina Kusumawardani ; Siti Oryza Khairunnisa</i> | |
| RESEARCH ON ANSWER SELECTION BASED ON LSTM | 357 |
| <i>Yangsen Zhang ; Yuanyuan Peng</i> | |
| SHARED REPRESENTATION LEARNING WITH SELF-ATTENTION FOR CROSS-DOMAIN CHINESE HEDGE CUE RECOGNITION | 362 |
| <i>Huiwei Zhou ; Shixian Ning ; Zhe Liu ; Zhuang Liu ; Chengkun Lang</i> | |
| SIMPLE VECTOR REPRESENTATIONS OF E-COMMERCE PRODUCTS | 368 |
| <i>Abe Vallerian Siswanto ; Lilian Tjong ; Yordan Saputra</i> | |
| TOPICAL-RELEVANCE DETECTION USING ATTENTION-BASED NEURAL NETWORK | 373 |
| <i>Xia Li ; Zhanyuan Yang ; Minping Chen ; Wenhe Feng</i> | |

| | |
|---|------------|
| RELEVANCE-BASED AUTOMATED ESSAY SCORING VIA HIERARCHICAL RECURRENT MODEL | 378 |
| <i>Minping Chen ; Xia Li</i> | |
| Author Index | |