

2015 International Conference on Asian Language Processing (IALP 2015)

**Suzhou, China
24 – 25 October 2015**



IEEE Catalog Number: CFP1544I-POD
ISBN: 978-1-4673-9597-7

**Copyright © 2015 by the Institute of Electrical and Electronic 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 publication is a representation of what appears in the IEEE Digital Libraries. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP1544I-POD
ISBN (Print-On-Demand):	978-1-4673-9597-7
ISBN (Online):	978-1-4673-9596-0

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

CURRAN ASSOCIATES INC.
proceedings
.com

Table of Contents

Invited Talk

Invited Talk1	NLP for Microblog Summarization <i>Kam Fai Wong</i>	I
Invited Talk2	Speech Recognition and Keyword Search for Low Resource Languages <i>Bin Ma</i>	II
Invited Talk3	Technologies behind Digital Personal Assistants <i>Mei-Yuh Hwang</i>	III

Machine Translation

0001-119	A Novel Method to Optimize Training Data for Translation Model Adaptation <i>Hao Liu, Yu Hong, Liang Yao, Le Liu, Jianmin Yao, Qiaoming Zhu</i>	1
0002-130	Improving Interactive Machine Translation via Multiple Positive Constraints <i>Na Ye, Yitao Fu, Guiping Zhang, Chuang Wu, Dongfeng Cai</i>	5
0003-136	Refining Parallel Quality for Machine Translation <i>Huimin Gong, Xiangyu Duan, Min Zhang</i>	9
0004-144	Bilingual Lexicon Extraction using Locally Weighted Linear Regression from Comparable Corpora <i>Chunyue Zhang, Tiejun Zhao</i>	13
0005-147	Machine Translation from Japanese and French to Vietnamese, the Difference among Language Families <i>Thi-Ngoc-Diep Do, Masao Utiyama, Eiichiro Sumita</i>	17
0006-155	Extracting Bilingual Multi-word Expressions for Low-resource Statistical Machine Translation <i>Linyu Wei, Miao Li, Lei Chen, Zhenxin Yang, Kai Sun, Man Yuan</i>	21
0007-157	Constructing Morpheme-based Translation Model for Mongolian-Chinese SMT <i>Zhenxin Yang, Miao Li, Lei Chen, Linyu Wei, Jinxing Wu, Sheng Chen</i>	25
0008-158	Context-extended Phrase Reordering Model for Pivot-based Statistical Machine Translation <i>Xiaoning Zhu, Tiejun Zhao, Yiming Cui, Conghui Zhu</i>	29
0009-161	Exploring Hybrid Character-Words Representational Unit in Classical-to-Modern Chinese Machine Translation <i>Hongyang Zhang, Muyun Yang, Tiejun Zhao</i>	33
0010-160	POS-based Word Alignment for Small Corpus <i>Jyoti Srivastava, Sudip Sanyal</i>	37

Word Segmentation and Tagging

0011-111	Joint Chinese Word Segmentation and Punctuation Prediction Using Deep Recurrent Neural Network for Social Media Data <i>Kui Wu, Xuancong Wang, Nina Zhou, AiTi Aw, Haizhou Li</i>	41
0012-112	Unsupervised Learning of Agglutinated Morphology Using Nested Pitman-Yor Process Based Morpheme Induction Algorithm <i>Arun Kumar, Luis Pedro, Antoni Oliver</i>	45
0013-127	What We Need is Word, Not Morpheme; Constructing Word Analyzer for Japanese <i>Kazuhide Yamamoto, Yuki Miyanishi, Kanji Takahashi, Yoshiki Inomata, Yuki Mikami, Yuta Sudo</i>	49

0014-129	Paninian Grammar Based Hindi Dialogue Anaphora Resolution <i>Vandan Mujadia, Darshan Agarwal, Radhika Mamidi, Dipti Misra Sharma</i>	53
0015-141	Automatic Segmentation of Chinese Mandarin Speech into Syllable-like <i>Jian Li, Furao Shen</i>	57

Resources

0016-122	Construction of Japanese Semantically Compatible Words Resource <i>Kazuhide Yamamoto, Kanji Takahashi</i>	61
0017-124	Japanese Sentence Compression using Simple English Wikipedia <i>Shunsuke Takeno, Kazuhide Yamamoto</i>	65
0018-146	The Annotation Study of Grammar Points for TCSL <i>Xiaoping Tan, Lijiao Yang</i>	69
0019-151	Attribute knowledge mining for Chinese word sense disambiguation <i>Jianyong Duan, Yao Fu, Xia Li</i>	73
0020-156	Extracting Coordinate Word Pairs for Dependency Parsing <i>Junjie Yu, Wenliang Chen</i>	78

Sentiment Analysis and Opinion Mining

0021-2	Recognizing Entailment in Chinese Texts with Feature Combination <i>Maofu Liu, Yifan Guo, Liqiang Nie</i>	82
0022-104	A Hybrid Method for Sentiment Classification in Chinese Movie Reviews Based on Sentiment Labels <i>Kai Zhao, Yaohong Jin</i>	86
0023-123	Multi-view Learning for Emotion Detection in Code-switching Texts <i>Sophia Yat Mei Lee, Zhongqing Wang</i>	90
0024-139	Learning Sentiment-inherent Word Embedding for Word-level and Sentence-Level Sentiment Analysis <i>Zhihua Zhang, Man Lan</i>	94
0025-163	The Expression of Singing Emotion - Contradicting the Constraints of Song <i>Gillian Chua, Qian Ci Chang, Ye Won Park, Paul Yaozhu Chan, Minghui Dong, Haizhou Li</i>	98

NLP Applications

0026-105	A Two-Step Approach for Event Factuality Identification <i>Zhong Qian, Peifeng Li, Qiaoming Zhu</i>	103
0027-108	Rumor Diffusion Purpose Analysis from Social Attribute to Social Content <i>Dazhen Lin, Yanping Lv, Donglin Cao</i>	107
0028-117	Correlation Analysis between Social Network Content and Query Intention <i>Jian Wang, Yu Hong, Kai Wang, Jianmin Yao, Qiaoming Zhu</i>	111
0029-137	An upgrading SentiVoice - a system for querying hotel service reviews via phone <i>Thien Khai Tran, Tuoi Thi Phan</i>	115

0030-143	Performance Scoring of Singing Voice	119
	<i>Yang Yu, Weisi Lin, Dong-Yan Huang, Minghui Dong, Haizhou Li</i>	

Text Processing

0031-103	Tibetan Text Classification Using Distributed Representations of Words	123
	<i>Tao Jiang, Hongzhi Yu, Bing Zhang</i>	
0032-121	Effect of Word Segmentation on Arabic Text Classification	127
	<i>Abdulmohsen Al-Thubaity, Abdullah Al-Subaie</i>	
0033-133	Weighted Document Frequency for Feature Selection in Text Classification	132
	<i>Baoli Li, Qiuling Yan, Zhenqiang Xu, Guicai Wang</i>	
0034-132	Towards Improving the Performance of Vector Space Model for Chinese Frequently Asked Question Answering	136
	<i>Ridong Jiang, Seokhwam Kim, Rafael E. Banchs, Haizhou Li</i>	
0035-114	Learning from the Past: Improving News Summarization with Past News Articles	140
	<i>Feng Li, Yan Chen, Zhoujun Li</i>	
0036-115	A Kanji Learning System Based on Automatic Question Sentence Generation	144
	<i>Jianwei Zhang, Jun Takuma</i>	
0037-120	A Comparative Study on Collectives of TermWeighting Methods for Extractive Presentation Speech Summarization	148
	<i>Jian Zhang, Huaqiang Yuan</i>	
0038-152	Document Summarization Based on Semantic Representations	152
	<i>Hui Zhang, Xueliang Zhang, Guanglai Gao</i>	
0039-131	Subtopic Boundary Identification in Hindi Dialogue	156
	<i>Darshan Agarwal, Vandan Mujadia, Radhika Mamidi</i>	

NER and Parsing

0040-1	Improving Person Name Recognition Quality in Chinese Text with Reinforced Processing of Ambiguities	160
	<i>Shuangyong Song, Zhongguang Zheng, Yao Meng</i>	
0041-145	Named Entity Recognizer for Less Resourced Language Kokborok	164
	<i>Braja Gopal Patra, Nuna Debbarma, Aby Abahai T., Dipankar Das, Sivaji Bandyopadhyay</i>	
0042-154	Mongolian Named Entity Recognition Using Suffixes Segmentation	169
	<i>Weihua Wang, Feilong Bao, Guanglai Gao</i>	
0043-150	Mandarin Prosodic Word Prediction using Dependency Relationships	173
	<i>Zhengchen Zhang, Fuxiang Wu, Minghui Dong, Fugen Zhou</i>	
0044-110	A Re-ranking Model for Dependency Parsing with Knowledge Graph Embeddings	177
	<i>A-Yeong Kim, Hyun-Je Song, Seong-Bae Park, Sang-Jo Lee</i>	

Information Extraction

0045-138	Toward Better Keywords Extraction <i>Shihua Xu, Fang Kong</i>	181
0046-142	LDA topic model for microblog recommendation <i>Jianyong Duan, Yamin Ai, Xia Li</i>	185
0047-153	Multilingual Viewpoint Detection from News Comments <i>Bei Shi, Lam Wai, Lidong Bingy, Yinqing Xu</i>	189
0048-109	Topic2Vec: Learning Distributed Representations of Topics <i>Liqiang Niu, Xinyu Dai, Jianbing Zhang, Jiajun Chen</i>	193
0049-148	Recognizing Nuclearity between Chinese Discourse Units <i>Xiaomin Chu, Zhongqing Wang, Qiaoming Zhu, Guodong Zhou</i>	197