

Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)

Industry Track

Singapore
6 - 10 December 2023

ISBN: 978-1-7138-8591-7

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2023) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2024)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

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
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>BeautifulPrompt: Towards Automatic Prompt Engineering for Text-to-Image Synthesis</i> Tingfeng Cao, Chengyu Wang, Bingyan Liu, Ziheng Wu, Jinhui Zhu and Jun Huang	1
<i>Enhancing Language Model with Unit Test Techniques for Efficient Regular Expression Generation</i> Chenhui Mao, Xiexiong Lin, Xin Jin and Xin Zhang	12
<i>A Comparative Analysis of Task-Agnostic Distillation Methods for Compressing Transformer Language Models</i> Takuma Udagawa, Aashka Trivedi, Michele Merler and Bishwaranjan Bhattacharjee	20
<i>Towards Effective Automatic Debt Collection with Persona Awareness</i> Tong Zhang, Junhong Liu, Chen Huang, Jia Liu, Hongru Liang, Zujie Wen and Wenqiang Lei	32
<i>Gatekeeper to save COGS and improve efficiency of Text Prediction</i> Nidhi Tiwari, Sneha Kola, Milos Milunovic, Si-qing Chen and Marjan Slavkovski	46
<i>Efficient Transformer Knowledge Distillation: A Performance Review</i> Nathan Brown, Ashton Williamson, Tahj Anderson and Logan Lawrence	54
<i>CDD: A Large Scale Dataset for Legal Intelligence Research</i> Changzhen Ji, Yating Zhang, Adam Jatowt and Haipang Wu	66
<i>MUST&P-SRL: Multi-lingual and Unified Syllabification in Text and Phonetic Domains for Speech Representation Learning</i> Noé Tits	74
<i>Personalized Dense Retrieval on Global Index for Voice-enabled Conversational Systems</i> Masha Belyi, Charlotte Dzialo, Chaitanya Dwivedi, Prajit Reddy Muppidi and Kanna Shimizu	83
<i>Text2Topic: Multi-Label Text Classification System for Efficient Topic Detection in User Generated Content with Zero-Shot Capabilities</i> Fengjun Wang, Moran Beladev, Ofri Kleinfeld, Elina Frayerman, Tal Shachar, Eran Fainman, Karen Lastmann Assaraf, Sarai Mizrachi and Benjamin Wang	93
<i>Deep Metric Learning to Hierarchically Rank - An Application in Product Retrieval</i> Kee Kiat Koo, Ashutosh Joshi, Nishaanth Reddy, Karim Bouyarmane, Ismail Tutar, Vaclav Petricek and Changhe Yuan	104
<i>A Pretrained Language Model for Cyber Threat Intelligence</i> Youngja Park and Weiqiu You	113
<i>SAMP: A Model Inference Toolkit of Post-Training Quantization for Text Processing via Self-Adaptive Mixed-Precision</i> Rong Tian, Zijng Zhao, Weijie Liu, Haoyan Liu, Weiquan Mao, Zhe Zhao and Kan Zhou	123
<i>KD-Boost: Boosting Real-Time Semantic Matching in E-commerce with Knowledge Distillation</i> Sanjay Agrawal, Vivek Sembium and Ankith M S	131
<i>Multi-teacher Distillation for Multilingual Spelling Correction</i> Jingfen Zhang, Xuan Guo, Sravan Bodapati and Christopher Potts	142
<i>Does Named Entity Recognition Truly Not Scale Up to Real-world Product Attribute Extraction?</i> Wei-Te Chen, Keiji Shinzato, Naoki Yoshinaga and Yandi Xia	152

<i>Investigating Table-to-Text Generation Capabilities of Large Language Models in Real-World Information Seeking Scenarios</i>	
Yilun Zhao, Haowei Zhang, Shengyun Si, Linyong Nan, Xiangru Tang and Arman Cohan . . .	160
<i>TMID: A Comprehensive Real-world Dataset for Trademark Infringement Detection in E-Commerce</i>	
Tongxin Hu, Zhuang Li, Xin Jin, Lizhen Qu and Xin Zhang	176
<i>Joint Dialogue Topic Segmentation and Categorization: A Case Study on Clinical Spoken Conversations</i>	
Zhengyuan Liu, Siti Umairah Md Salleh, Hong Choon Oh, Pavitra Krishnaswamy and Nancy Chen	185
<i>AdapterDistillation: Non-Destructive Task Composition with Knowledge Distillation</i>	
Junjie Wang, Yicheng Chen, Wangshu Zhang, Sen Hu, Teng Xu and Jing Zheng	194
<i>PROMINET: Prototype-based Multi-View Network for Interpretable Email Response Prediction</i>	
Yuqing Wang, Prashanth Vijayaraghavan and Ehsan Degan	202
<i>Retrieval-Enhanced Dual Encoder Training for Product Matching</i>	
Justin Chiu	216
<i>WordArt Designer: User-Driven Artistic Typography Synthesis using Large Language Models</i>	
Jun-Yan He, Zhi-Qi Cheng, Chenyang Li, Jingdong Sun, Wangmeng Xiang, Xianhui Lin, Xiaoyang Kang, Zengke Jin, Yusen Hu, Bin Luo, Yifeng Geng and Xuansong Xie	223
<i>Lattice Path Edit Distance: A Romanization-aware Edit Distance for Extracting Misspelling-Correction Pairs from Japanese Search Query Logs</i>	
Nobuhiro Kaji	233
<i>Learning Multilingual Sentence Representations with Cross-lingual Consistency Regularization</i>	
Pengzhi Gao, Liwen Zhang, Zhongjun He, Hua Wu and Haifeng Wang	243
<i>Unveiling Identity Biases in Toxicity Detection : A Game-Focused Dataset and Reactivity Analysis Approach</i>	
Josiane Van Dorpe, Zachary Yang, Nicolas Grenon-Godbout and Grégoire Winterstein	263
<i>ORANGE: Text-video Retrieval via Watch-time-aware Heterogeneous Graph Contrastive Learning</i>	
Yucheng Lin, Tim Chang, Yaning Chang, Jianqiang Ma, Donghui Li, Ting Peng, Zang Li, Zhiyi Zhou and Feng Wang	275
<i>Compute-Efficient Churn Reduction for Conversational Agents</i>	
Christopher Hidey and Sarthak Sarthak	284
<i>Empower Large Language Model to Perform Better on Industrial Domain-Specific Question Answering</i>	
Fangkai Yang, Pu Zhao, Zezhong Wang, Lu Wang, Bo Qiao, Jue Zhang, Mohit Garg, Qingwei Lin, Saravan Rajmohan and Dongmei Zhang	294
<i>Enhancing Extreme Multi-Label Text Classification: Addressing Challenges in Model, Data, and Evaluation</i>	
Dan Li, Zi Long Zhu, Janneke van de Loo, Agnes Masip Gomez, Vikrant Yadav, Georgios Tsatsaronis and Zubair Afzal	313
<i>Query-aware Multi-modal based Ranking Relevance in Video Search</i>	
Chengcan Ye, Ting Peng, Tim Chang, Zhiyi Zhou and Feng Wang	322
<i>Coordinated Replay Sample Selection for Continual Federated Learning</i>	
Jack Good, Jimit Majmudar, Christophe Dupuy, Jixuan Wang, Charith Peris, Clement Chung, Richard Zemel and Rahul Gupta	331

<i>Building Real-World Meeting Summarization Systems using Large Language Models: A Practical Perspective</i>	
Md Tahmid Rahman Laskar, Xue-Yong Fu, Cheng Chen and Shashi Bhushan TN	343
<i>Creator Context for Tweet Recommendation</i>	
Spurthi Amba Hombaiah, Tao Chen, Mingyang Zhang, Michael Bendersky, Marc Najork, Matt Colen, Sergey Levi, Vladimir Ofitserov and Tanvir Amin	353
<i>AdaBERT-CTC: Leveraging BERT-CTC for Text-Only Domain Adaptation in ASR</i>	
Tyler Vuong, Karel Mundnich, Dhanush Bekal, Veera Raghavendra Elluru, Srikanth Ronanki and Sravan Bodapati	364
<i>Conversing with databases: Practical Natural Language Querying</i>	
Denis Kochedykov, Fenglin Yin and Sreevidya Khatravath	372
<i>AART: AI-Assisted Red-Teaming with Diverse Data Generation for New LLM-powered Applications</i>	
Bhaktipriya Radharapu, Kevin Robinson, Lora Aroyo and Preethi Lahoti	380
<i>Speakerly: A Voice-based Writing Assistant for Text Composition</i>	
Dhruv Kumar, Vipul Raheja, Alice Kaiser-Schatzlein, Robyn Perry, Apurva Joshi, Justin Hugues-Nuger, Samuel Lou and Navid Chowdhury	396
<i>Are ChatGPT and GPT-4 General-Purpose Solvers for Financial Text Analytics? A Study on Several Typical Tasks</i>	
Xianzhi Li, Samuel Chan, Xiaodan Zhu, Yulong Pei, Zhiqiang Ma, Xiaomo Liu and Sameena Shah	408
<i>CL-QR: Cross-Lingual Enhanced Query Reformulation for Multi-lingual Conversational AI Agents</i>	
Zhongkai Sun, Zhengyang Zhao, Sixing Lu, Chengyuan Ma, Xiaohu Liu, Xing Fan, Wei Shen and Chenlei Guo	423
<i>Improving Contextual Query Rewrite for Conversational AI Agents through User-preference Feedback Learning</i>	
Zhongkai Sun, Yingxue Zhou, Jie Hao, Xing Fan, Yanbin Lu, Chengyuan Ma, Wei Shen and Chenlei Guo	432
<i>Scaling Neural ITN for Numbers and Temporal Expressions in Tamil: Findings for an Agglutinative Low-resource Language</i>	
Bhavuk Singhal, Sindhuja Gopalan, Amrith Krishna and Malolan Chetlur	440
<i>EELBERT: Tiny Models through Dynamic Embeddings</i>	
Gabrielle Cohn, Rishika Agarwal, Deepanshu Gupta and Siddharth Patwardhan	451
<i>Gold Standard Bangla OCR Dataset: An In-Depth Look at Data Preprocessing and Annotation Processes</i>	
Hasmot Ali, AKM Shahariar Azad Rabby, Md Majedul Islam, A.K.M Mahamud, Nazmul Hasan and Fuad Rahman	460
<i>PILLOW: Enhancing Efficient Instruction Fine-tuning via Prompt Matching</i>	
Zhenting Qi, Xiaoyu Tan, Shaojie Shi, Chao Qu, Yinghui Xu and Yuan Qi	471
<i>Welcome to the Real World: Efficient, Incremental and Scalable Key Point Analysis</i>	
Lilach Eden, Yoav Kantor, Matan Orbach, Yoav Katz, Noam Slonim and Roy Bar-Haim	483
<i>Automatic Linking of Judgements to UK Supreme Court Hearings</i>	
Hadeel Saadany and Constantin Orasan	492

<i>Automatic Marketing Theme and Commodity Construction System for E-commerce</i>	
Zhiping Wang, Peng Lin, Hainan Zhang, Hongshen Chen, Tianhao Li, Zhuoye Ding, Sulong Xu and Jinghe Hu	501
<i>Towards Safer Operations: An Expert-involved Dataset of High-Pressure Gas Incidents for Preventing Future Failures</i>	
Shumpei Inoue, Minh-Tien Nguyen, Hiroki Mizokuchi, Tuan-Anh D. Nguyen, Huu-Hiep Nguyen and Dung Le	509
<i>An Auxiliary Task Boosted Multi-task Learning Method for Service Account Retrieval with Limited Human Annotation</i>	
Yuanzhou Yao, Zhao Zhang, Kaijia Yang, Huasheng Liang, Qiang Yan and Yongjun Xu	522
<i>VKIE: The Application of Key Information Extraction on Video Text</i>	
Siyu An, Ye Liu, Haoyuan Peng and Di Yin	532
<i>Investigating the Role and Impact of Disfluency on Summarization</i>	
Varun Nathan, Ayush Kumar and Jithendra Vepa	541
<i>InsightNet : Structured Insight Mining from Customer Feedback</i>	
Sandeep Sricharan Mukku, Manan Soni, Chetan Aggarwal, Jitenkumar Rana, Promod Yenigalla, Rashmi Patange and Shyam Mohan	552
<i>E2E Spoken Entity Extraction for Virtual Agents</i>	
Karan Singla, Yeon-Jun Kim and Srinivas Bangalore	567
<i>Generative Models for Product Attribute Extraction</i>	
Ansel Blume, Nasser Zalmout, Heng Ji and Xian Li	575
<i>CarExpert: Leveraging Large Language Models for In-Car Conversational Question Answering</i>	
Md Rashad Al Hasan Rony, Christian Suess, Sinchana Ramakanth Bhat, Viju Sudhi, Julia Schneider, Maximilian Vogel, Roman Teucher, Ken E. Friedl and Soumya Sahoo	586
<i>BUSTER: a BUSIness Transaction Entity Recognitiondataset</i>	
Andrea Zugarini, Andrew Zamai, Marco Ernandes and Leonardo Rigutini	605
<i>Multi-word Tokenization for Sequence Compression</i>	
Leonidas Gee, Leonardo Rigutini, Marco Ernandes and Andrea Zugarini	612
<i>JarviX: A LLM No code Platform for Tabular Data Analysis and Optimization</i>	
Shang-Ching Liu, ShengKun Wang, Tsungyao Chang, Wenqi Lin, Chung-Wei Hsiung, Yi-Chen Hsieh, Yu-Ping Cheng, Sian-Hong Luo and Jianwei Zhang	622
<i>Retrieve and Copy: Scaling ASR Personalization to Large Catalogs</i>	
Sai Muralidhar Jayanthi, Devang Kulshreshtha, Saket Dingliwal, Srikanth Ronanki and Sravan Bodapati	631
<i>STEER: Semantic Turn Extension-Expansion Recognition for Voice Assistants</i>	
Leon Zhang, Jiarui Lu, Joel Ruben Antony Moniz, Aditya Kulkarni, Dhivya Piraviperumal, Tien Dung Tran, Nick Tzou and Hong Yu	640
<i>Self-Criticism: Aligning Large Language Models with their Understanding of Helpfulness, Honesty, and Harmlessness</i>	
Xiaoyu Tan, Shaojie Shi, Xihe Qiu, Chao Qu, Zhenting Qi, Yinghui Xu and Yuan Qi	650
<i>InstructPTS: Instruction-Tuning LLMs for Product Title Summarization</i>	
Besnik Fetahu, Zhiyu Chen, Oleg Rokhlenko and Shervin Malmasi	663

<i>LLM4Vis: Explainable Visualization Recommendation using ChatGPT</i>	
Lei Wang, Songheng Zhang, Yun Wang, Ee-Peng Lim and Yong Wang	675
<i>DUBLIN: Visual Document Understanding By Language-Image Network</i>	
Kriti Aggarwal, Aditi Khandelwal, Kumar Tanmay, Owais Khan Mohammed, Qiang Liu, Monojit Choudhury, Hardik Hansrajbhai Chauhan, Subhojit Som, Vishrav Chaudhary and Saurabh Tiwary	693
<i>DocumentNet: Bridging the Data Gap in Document Pre-training</i>	
Lijun Yu, Jin Miao, Xiaoyu Sun, Jiayi Chen, Alexander Hauptmann, Hanjun Dai and Wei Wei	707
<i>Relevance-assisted Generation for Robust Zero-shot Retrieval</i>	
Jihyuk Kim, Minsoo Kim, Joonsuk Park and Seung-won Hwang	723
<i>Too much of product information : Don't worry, let's look for evidence!</i>	
Aryan Jain, Jitenkumar Rana and Chetan Aggarwal	732
<i>Harnessing LLMs for Temporal Data - A Study on Explainable Financial Time Series Forecasting</i>	
Xinli Yu, Zheng Chen and Yanbin Lu	739
<i>ViGPTQA - State-of-the-Art LLMs for Vietnamese Question Answering: System Overview, Core Models Training, and Evaluations</i>	
Minh Thuan Nguyen, Khanh Tung Tran, Nhu Van Nguyen and Xuan-Son Vu	754
<i>An Integrated Search System for Korea Weather Data</i>	
Jinkyung Jo, Dayeon Ki, Soyoung Yoon and Minjoon Seo	765
<i>Adaptive Hyper-parameter Learning for Deep Semantic Retrieval</i>	
Mingming Li, Chunyuan Yuan, Huimu Wang, Peng Wang, Jingwei Zhuo, Binbin Wang, Lin Liu and Sulong Xu	775
<i>On Sample-Efficient Code Generation</i>	
Hojae Han, Yu Jin Kim, Byoungjip Kim, Youngwon Lee, Kyungjae Lee, Kyungmin Lee, Moontae Lee, Kyunghoon Bae and Seung-won Hwang	783
<i>Batch Prompting: Efficient Inference with Large Language Model APIs</i>	
Zhoujun Cheng, Jungo Kasai and Tao Yu	792
<i>Graph Meets LLM: A Novel Approach to Collaborative Filtering for Robust Conversational Understanding</i>	
Zheng Chen, Ziyang Jiang, Fan Yang, Eunah Cho, Xing Fan, Xiaojiang Huang, Yanbin Lu and Aram Galstyan	811
<i>DELPHI: Data for Evaluating LLMs' Performance in Handling Controversial Issues</i>	
David Sun, Artem Abzaliev, Hadas Kotek, Christopher Klein, Zidi Xiu and Jason D Williams	820
<i>Angel: Enterprise Search System for the Non-Profit Industry</i>	
Saiful Haq, Ashutosh Sharma and Pushpak Bhattacharyya	828