

61st Annual Meeting of the Association for Computational Linguistics (ACL 2023)

Industry Track

Toronto, Canada
10 - 12 July 2023

ISBN: 978-1-7138-8177-3

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>CWSeg: An Efficient and General Approach to Chinese Word Segmentation</i> Dedong Li, Rui Zhao and Fei Tan	1
<i>Knowledge is Power": Constructing Knowledge Graph of Abdominal Organs and Using Them for Automatic Radiology Report Generation</i> Kaveri Kale, Pushpak Bhattacharyya, Aditya Shetty, Milind Gune, Kush Shrivastava, Rustom Lawyer and Spruha Biswas	11
<i>Hunt for Buried Treasures: Extracting Unclaimed Embodiments from Patent Specifications</i> Chikara Hashimoto, Gautam Kumar, Shuichiro Hashimoto and Jun Suzuki	25
<i>MathPrompter: Mathematical Reasoning using Large Language Models</i> Shima Imani, Liang Du and Harsh Shrivastava	37
<i>Constrained Policy Optimization for Controlled Self-Learning in Conversational AI Systems</i> Mohammad Kachuee and Sungjin Lee	43
<i>pNLP-Mixer: an Efficient all-MLP Architecture for Language</i> Francesco Fusco, Damian Pascual, Peter Staar and Diego Antognini	53
<i>Extracting Text Representations for Terms and Phrases in Technical Domains</i> Francesco Fusco and Diego Antognini	61
<i>CocaCLIP: Exploring Distillation of Fully-Connected Knowledge Interaction Graph for Lightweight Text-Image Retrieval</i> Jiapeng Wang, Chengyu Wang, Xiaodan Wang, Jun Huang and Lianwen Jin	71
<i>KG-FLIP: Knowledge-guided Fashion-domain Language-Image Pre-training for E-commerce</i> Qinjin Jia, Yang Liu, Daoping Wu, Shaoyuan Xu, Huidong Liu, Jinmiao Fu, Roland Vollgraf and Bryan Wang	81
<i>Domain-specific transformer models for query translation</i> Mandar Kulkarni, Nikesh Garera and Anusua Trivedi	89
<i>Label efficient semi-supervised conversational intent classification</i> Mandar Kulkarni, Kyung Kim, Nikesh Garera and Anusua Trivedi	96
<i>xPQA: Cross-Lingual Product Question Answering in 12 Languages</i> Xiaoyu Shen, Akari Asai, Bill Byrne and Adria De Gispert	103
<i>Learn over Past, Evolve for Future: Forecasting Temporal Trends for Fake News Detection</i> Beizhe Hu, Qiang Sheng, Juan Cao, Yongchun Zhu, Danding Wang, Zhengjia Wang and Zhiwei Jin	116
<i>AVEN-GR: Attribute Value Extraction and Normalization using product GRaphs</i> Thomas Ricatte and Donato Crisostomi	126
<i>GKD: A General Knowledge Distillation Framework for Large-scale Pre-trained Language Model</i> Shicheng Tan, Weng Lam Tam, Yuanchun Wang, Wenwen Gong, Shu Zhao, Peng Zhang and Jie Tang	134
<i>FashionKLIP: Enhancing E-Commerce Image-Text Retrieval with Fashion Multi-Modal Conceptual Knowledge Graph</i> Xiaodan Wang, Chengyu Wang, Lei Li, Zhixu Li, Ben Chen, Linbo Jin, Jun Huang, Yanghua Xiao and Ming Gao	149

<i>Entity Contrastive Learning in a Large-Scale Virtual Assistant System</i> Jonathan Rubin, Jason Crowley, George Leung, Morteza Ziyadi and Maria Minakova	159
<i>Tab-Cleaner: Weakly Supervised Tabular Data Cleaning via Pre-training for E-commerce Catalog</i> Kewei Cheng, Xian Li, Zhengyang Wang, Chenwei Zhang, Binxuan Huang, Yifan Ethan Xu, Xin Luna Dong and Yizhou Sun	172
<i>Toward More Accurate and Generalizable Evaluation Metrics for Task-Oriented Dialogs</i> Abishek Komma, Nagesh Panyam Chandrasekarastry, Timothy Leffel, Anuj Goyal, Angeliki Metallinou, Spyros Matsoukas and Aram Galstyan	186
<i>Tab-CQA: A Tabular Conversational Question Answering Dataset on Financial Reports</i> Chuang Liu, Junzhuo Li and Deyi Xiong	196
<i>KoSBI: A Dataset for Mitigating Social Bias Risks Towards Safer Large Language Model Applications</i> Hwaran Lee, Seokhee Hong, Joonsuk Park, Takyoung Kim, Gunhee Kim and Jung-woo Ha .	208
<i>Improving Knowledge Production Efficiency With Question Answering on Conversation</i> Changlin Yang, Siye Liu, Sen Hu, Wangshu Zhang, Teng Xu and Jing Zheng	225
<i>Mitigating the Burden of Redundant Datasets via Batch-Wise Unique Samples and Frequency-Aware Losses</i> Donato Crisostomi, Andrea Caciolai, Alessandro Pedrani, Kay Rottmann, Alessandro Manzotti, Enrico Palumbo and Davide Bernardi	235
<i>Distilled Language Models are economically efficient for the enterprise. ...mostly.</i> Kristen Howell, Gwen Christian, Pavel Fomitchov, Gitit Kehat, Julianne Marzulla, Leanne Rolston, Jadin Tredup, Ilana Zimmerman, Ethan Selfridge and Joseph Bradley	248
<i>Application-Agnostic Language Modeling for On-Device ASR</i> Markus Nussbaum-thom, Lyan Verwimp and Youssef Oualil	268
<i>Building Accurate Low Latency ASR for Streaming Voice Search in E-commerce</i> Abhinav Goyal and Nikesh Garera	276
<i>PLAtE: A Large-scale Dataset for List Page Web Extraction</i> Aidan San, Yuan Zhuang, Jan Bakus, Colin Lockard, David Ciemiewicz, Sandeep Atluri, Kevin Small, Yangfeng Ji and Heba Elfardy	284
<i>Rapid Diffusion: Building Domain-Specific Text-to-Image Synthesizers with Fast Inference Speed</i> Bingyan Liu, Weifeng Lin, Zhongjie Duan, Chengyu Wang, Wu Ziheng, Zhang Zipeng, Kui Jia, Lianwen Jin, Cen Chen and Jun Huang	295
<i>Large Scale Generative Multimodal Attribute Extraction for E-commerce Attributes</i> Anant Khandelwal, Happy Mittal, Shreyas Kulkarni and Deepak Gupta	305
<i>Consistent Text Categorization using Data Augmentation in e-Commerce</i> Noa Avigdor, Guy Horowitz, Ariel Raviv and Stav Yanovsky Daye	313
<i>An efficient method for Natural Language Querying on Structured Data</i> Hanoz Bhatena, Aviral Joshi and Prateek Singh	322
<i>Boosting Transformers and Language Models for Clinical Prediction in Immunotherapy</i> Zekai Chen, Mariann Micsinai Balan and Kevin Brown	332
<i>EvolveMT: an Ensemble MT Engine Improving Itself with Usage Only</i> Kamer Yüksel, Ahmet Gunduz, Mohamed Al-badrashiny and Hassan Sawaf	341

<i>A Static Evaluation of Code Completion by Large Language Models</i>	
Hantian Ding, Varun Kumar, Yuchen Tian, Zijian Wang, Rob Kwiatkowski, Xiaopeng Li, Murali Krishna Ramanathan, Baishakhi Ray, Parminder Bhatia and Sudipta Sengupta	347
<i>Scalable and Safe Remediation of Defective Actions in Self-Learning Conversational Systems</i>	
Sarthak Ahuja, Mohammad Kachuee, Fatemeh Sheikholeslami, Weiqing Liu and Jaeyoung Do	361
<i>MobileNMT: Enabling Translation in 15MB and 30ms</i>	
Ye Lin, Xiaohui Wang, Zhexi Zhang, Mingxuan Wang, Tong Xiao and Jingbo Zhu	368
<i>Multi-doc Hybrid Summarization via Salient Representation Learning</i>	
Min Xiao	379
<i>SaFER: A Robust and Efficient Framework for Fine-tuning BERT-based Classifier with Noisy Labels</i>	
Zhenting Qi, Xiaoyu Tan, Chao Qu, Yinghui Xu and Yuan Qi	390
<i>Chemical Language Understanding Benchmark</i>	
Yunsoo Kim, Hyuk Ko, Jane Lee, Hyun Young Heo, Jinyoung Yang, Sungsoo Lee and Kyu-hwang Lee	404
<i>HyperT5: Towards Compute-Efficient Korean Language Modeling</i>	
Dongju Park, Soonwon Ka, Kang Min Yoo, Gichang Lee and Jaewook Kang	412
<i>Semantic Ambiguity Detection in Sentence Classification using Task-Specific Embeddings</i>	
Jong Myoung Kim, Young-jun Lee, Sangkeun Jung and Ho-jin Choi	425
<i>Reliable and Interpretable Drift Detection in Streams of Short Texts</i>	
Ella Rabinovich, Matan Vetzler, Samuel Ackerman and Ateret Anaby Tavor	438
<i>Sharing Encoder Representations across Languages, Domains and Tasks in Large-Scale Spoken Language Understanding</i>	
Jonathan Hueser, Judith Gaspers, Thomas Gueudre, Chandana Prakash, Jin Cao, Daniil Sorokin, Quynh Do, Nicolas Anastassacos, Tobias Falke and Turan Gojayev	447
<i>Annotating Research Infrastructure in Scientific Papers: An NLP-driven Approach</i>	
Seyed Amin Tabatabaei, Georgios Cheirmpas, Marius Doornenbal, Alberto Zigoni, Veronique Moore and Georgios Tsatsaronis	457
<i>Event-Centric Query Expansion in Web Search</i>	
Yanan Zhang, Weijie Cui, Yangfan Zhang, Xiaoling Bai, Zhe Zhang, Jin Ma, Xiang Chen and Tianhua Zhou	464
<i>Transferable and Efficient: Unifying Dynamic Multi-Domain Product Categorization</i>	
Shansan Gong, Zelin Zhou, Shuo Wang, Fengjiao Chen, Xiujie Song, Xuezhi Cao, Yunsen Xian and Kenny Zhu	476
<i>DISCOSQA: A Knowledge Base Question Answering System for Space Debris based on Program Induction</i>	
Paul Darm, Antonio Valerio Miceli Barone, Shay B. Cohen and Annalisa Riccardi	487
<i>BADGE: Speeding Up BERT Inference after Deployment via Block-wise Bypasses and Divergence-based Early Exiting</i>	
Wei Zhu, Peng Wang, Yuan Ni, Guotong Xie and Xiaoling Wang	500
<i>K-pop and fake facts: from texts to smart alerting for maritime security</i>	
Maxime Prieur, Souhir Gahbiche, Guillaume Gadek, Sylvain Gatepaille, Kilian Vasnier and Valerian Justine	510

<i>Evaluating Embedding APIs for Information Retrieval</i>	
Ehsan Kamaloo, Xinyu Zhang, Odunayo Ogundepo, Nandan Thakur, David Alfonso-hermelo, Mehdi Rezagholizadeh and Jimmy Lin	518
<i>Domain-Agnostic Neural Architecture for Class Incremental Continual Learning in Document Processing Platform</i>	
Mateusz Wójcik, Witold Kościukiewicz, Mateusz Baran, Tomasz Kajdanowicz and Adam Gonczarek	527
<i>Regression-Free Model Updates for Spoken Language Understanding</i>	
Andrea Caciolai, Verena Weber, Tobias Falke, Alessandro Pedrani and Davide Bernardi	538
<i>Reducing cohort bias in natural language understanding systems with targeted self-training scheme</i>	
Dieu-thu Le, Gabriela Hernandez, Bei Chen and Melanie Bradford	552
<i>Content Moderation for Evolving Policies using Binary Question Answering</i>	
Sankha Subhra Mullick, Mohan Bhambhani, Suhit Sinha, Akshat Mathur, Somya Gupta and Jidnya Shah	561
<i>Weighted Contrastive Learning With False Negative Control to Help Long-tailed Product Classification</i>	
Tianqi Wang, Lei Chen, Xiaodan Zhu, Younghun Lee and Jing Gao	574
<i>Towards Building a Robust Toxicity Predictor</i>	
Dmitriy Bessalov, Sourav Bhabesh, Yi Xiang, Liutong Zhou and Yanjun Qi	581
<i>AI Coach Assist: An Automated Approach for Call Recommendation in Contact Centers for Agent Coaching</i>	
Md Tahmid Rahman Laskar, Cheng Chen, Xue-yong Fu, Mahsa Azizi, Shashi Bhushan and Simon Corston-oliver	599
<i>Unified Contextual Query Rewriting</i>	
Yingxue Zhou, Jie Hao, Mukund Rungta, Yang Liu, Eunah Cho, Xing Fan, Yanbin Lu, Vishal Vasudevan, Kellen Gillespie and Zeynab Raesy	608
<i>Context-Aware Query Rewriting for Improving Users' Search Experience on E-commerce Websites</i>	
Simiao Zuo, Qingyu Yin, Haoming Jiang, Shaohui Xi, Bing Yin, Chao Zhang and Tuo Zhao	616
<i>Federated Learning of Gboard Language Models with Differential Privacy</i>	
Zheng Xu, Yanxiang Zhang, Galen Andrew, Christopher Choquette, Peter Kairouz, Brendan McMahan, Jesse Rosenstock and Yuanbo Zhang	629
<i>RadLing: Towards Efficient Radiology Report Understanding</i>	
Rikhiya Ghosh, Oladimeji Farri, Sanjeev Kumar Karn, Manuela Danu, Ramya Vunikili and Larisa Micu	640
<i>Predicting Customer Satisfaction with Soft Labels for Ordinal Classification</i>	
Etienne Manderscheid and Matthias Lee	652
<i>Accurate Training of Web-based Question Answering Systems with Feedback from Ranked Users</i>	
Liang Wang, Ivano Lauriola and Alessandro Moschitti	660
<i>SPM: A Split-Parsing Method for Joint Multi-Intent Detection and Slot Filling</i>	
Sheng Jiang, Su Zhu, Ruisheng Cao, Qingliang Miao and Kai Yu	668
<i>NAG-NER: a Unified Non-Autoregressive Generation Framework for Various NER Tasks</i>	
Xinpeng Zhang, Ming Tan, Jingfan Zhang and Wei Zhu	676

<i>Search Query Spell Correction with Weak Supervision in E-commerce</i>	
Vishal Kakkar, Chinmay Sharma, Madhura Pande and Surender Kumar	687
<i>Let's not Quote out of Context": Unified Vision-Language Pretraining for Context Assisted Image Captioning</i>	
Abisek Rajakumar Kalarani, Pushpak Bhattacharyya, Niyati Chhaya and Sumit Shekhar	695
<i>What, When, and How to Ground: Designing User Persona-Aware Conversational Agents for Engaging Dialogue</i>	
Deuksin Kwon, Sunwoo Lee, Ki Hyun Kim, Seojin Lee, Taeyoon Kim and Eric Davis	707
<i>CUPID: Curriculum Learning Based Real-Time Prediction using Distillation</i>	
Arindam Bhattacharya, Ankith Ms, Ankit Gandhi, Vijay Huddar, Atul Saroop and Rahul Bhagat	720
<i>Answering Unanswered Questions through Semantic Reformulations in Spoken QA</i>	
Pedro Faustini, Zhiyu Chen, Besnik Fetahu, Oleg Rokhlenko and Shervin Malmasi	729
<i>Exploring Zero and Few-shot Techniques for Intent Classification</i>	
Soham Parikh, Mitul Tiwari, Prashil Tumbade and Quaizar Vohra	744
<i>Referring to Screen Texts with Voice Assistants</i>	
Shruti Bhargava, Anand Dhoot, Ing-marie Jonsson, Hoang Long Nguyen, Alkesh Patel, Hong Yu and Vincent Renkens	752
<i>Generate-then-Retrieve: Intent-Aware FAQ Retrieval in Product Search</i>	
Zhiyu Chen, Jason Choi, Besnik Fetahu, Oleg Rokhlenko and Shervin Malmasi	763
<i>KAFA: Rethinking Image Ad Understanding with Knowledge-Augmented Feature Adaptation of Vision-Language Models</i>	
Zhiwei Jia, Pradyumna Narayana, Arjun Akula, Garima Pruthi, Hao Su, Sugato Basu and Varun Jampani	772
<i>Weakly supervised hierarchical multi-task classification of customer questions</i>	
Jitenkumar Rana, Promod Yenigalla, Chetan Aggarwal, Sandeep Sricharan Mukku, Manan Soni and Rashmi Patange	786
<i>Automated Digitization of Unstructured Medical Prescriptions</i>	
Megha Sharma, Tushar Vatsal, Srujana Merugu and Aruna Rajan	794