Soyeong Jeong

Updated May 16, 2024

Email: starsuzi@kaist.ac.kr **Homepage**: https://starsuzi.github.io **Google Scholar**: /Soyeong Jeong

Research Interest

My research interests are mainly on Retrieval-Augmented Generation (RAG) for solving open-domain language tasks and interpretation of large language models to enhance their interpretability in real-world applications. Not limited to, I am interested in broad topics on natural language understanding.

Education

KAIST Daejeon, Korea
Ph.D. in School of Computing Mar 2022 – Present
M.S. in School of Computing Mar 2020 – Feb 2022
Thesis: Information Retrieval by Augmenting Document Representation

Korea University Seoul, Korea
B.S. in Computer Science and Engineering (Graduated with Honor) Mar 2016 – Feb 2020
B.E. in Software Technology and Enterprise Program (Interdisciplinary Program)

Anyang Foreign Language High School Anyang, Korea
Prestigious high school for talented students (Major in English) Mar 2013 – Feb 2016

Employment

Graduate student, KAIST (Advisor: Prof. Jong Cheol Park) Mar 2020 - Present
- Conducted research on LLM-powered Retrieval-Augmented Generation (RAG) systems.
- Conducted research on the open conversational question-answering task.
- Conducted research on augmentation of both sparse and dense retrievers.

Undergrad. Research Assistant, Korea University (Advisor: Prof. Jaewoo Kang) Mar 2019 - Feb 2020
- Participated in the major recommendation project by embedding curriculum vectors.
- Participated in the food ingredient & drug graph construction project.

Research Intern, Seoul SW-SoC Convergence R&BD Center, ETRI Jul 2019 - Aug 2019
- Participated in the Speech Emotion Recognition project.

Research Intern, Artificial Intelligence Research Laboratory, ETRI Jan 2019 - Feb 2019
- Participated in the AIR project (Developing Artificial Social Intelligence for Human-Care Robots).

Publications

International Publications

[21] Typos that Broke the RAG's Back: Genetic Attack on RAG Pipeline by Simulating Documents in the Wild via Low-level Perturbations
Sukmin Cho, Soyeong Jeong, Jeongyeon Seo, Taeho Hwang, and Jong C. Park
Under review

[20] Self-Knowledge Distillation for Learning Ambiguity
Hancheol Park, Soyeong Jeong, Sukmin Cho, and Jong C. Park
Under review

[19] Ask LLMs Directly, "What shapes your bias?": Measuring Social Bias in Large Language Models
Jisu Shin, Hoyun Song, Huije Lee, Soyeong Jeong, and Jong C. Park
Findings of the Association for Computational Linguistics **(Findings of ACL)**, 2024

[18] Adaptive-RAG: Learning to Adapt Retrieval-Augmented Large Language Models through Question Complexity
Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang, and Jong C. Park
North American Chapter of the Association for Computational Linguistics **(NAACL)**, 2024

[17] Test-Time Self-Adaptive Small Language Models for Question Answering
Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang, and Jong C. Park
Findings of the Empirical Methods in Natural Language Processing **(Findings of EMNLP)**, 2023

[16] Knowledge-Augmented Language Model Verification
Jinheon Baek, Soyeong Jeong, Minki Kang, Jong C. Park, and Sung Ju Hwang
Empirical Methods in Natural Language Processing **(EMNLP)**, 2023

[15] Improving Zero-shot Reader by Reducing Distractions
from Irrelevant Documents in Open-Domain Question Answering
Sukmin Cho, Jeongyeon Seo, Soyeong Jeong, Jong C. Park
Findings of the Empirical Methods in Natural Language Processing **(Findings of EMNLP)**, 2023

[14] Phrase Retrieval for Open-Domain Conversational Question Answering
with Conversational Dependency Modeling via Contrastive Learning
Soyeong Jeong, Jinheon Baek, Sung Ju Hwang, and Jong C. Park
Findings of the Association for Computational Linguistics **(Findings of ACL)**, 2023

[13] Discrete Prompt Optimization via Constrained Generation for Zero-shot Re-ranker
Sukmin Cho, Soyeong Jeong, Jeong yeon Seo, Jong C. Park
Findings of Association for Computational Linguistics **(Findings of ACL)**, 2023.

[12] Realistic Conversational Question Answering with Answer Selection
based on Calibrated Confidence and Uncertainty Measurement
Soyeong Jeong, Jinheon Baek, Sung Ju Hwang, and Jong C. Park
Conference of the European Chapter of the Association for Computational Linguistics **(EACL)**, 2023

[11] Augmenting Document Representations for Dense Retrieval with Interpolation and Perturbation
Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang, and Jong C. Park
Annual Meeting of the Association for Computational Linguistics **(ACL)**, 2022 **(Oral)**

[10] Query Generation with External Knowledge for Dense Retrieval
Sukmin Cho, Soyeong Jeong, Wonsuk Yang, Jong C. Park
Deep Learning Inside Out at Association for Computational Linguistics **(DeeLIO @ ACL)**, 2022.

[9] Unsupervised Document Expansion for Information Retrieval with Stochastic Text Generation
Soyeong Jeong, Jinheon Baek, ChaeHun Park, and Jong C. Park
Scholarly Document Processing Workshop at NAACL **(SDP @ NAACL)**, 2021 **(Oral)**

[8] Development of Speech Emotion Recognition Algorithm using MFCC and Prosody
Hyejin Koo, Soyeong Jeong, Sungjae Yoon, Wonjong Kim
International Conference on Electronics, Information, and Communication **(ICEIC)**, 2020.

Domestic Publications, mostly written in Korean

[7] Controllable prompt tuning with relation dependent tokens
Jinseok Kim, Sukmin Cho, Soyeong Jeong, and Jong C. Park
Korea Computer Congress (KCC), 2023.

[6] Stopwords Mask Pooling for Dense Retrieval in Medical Domain
Dongho Choi, Hoyun Song, Soyeong Jeong, Sukmin Cho, and Jong C. Park
Korea Computer Congress (KCC), 2022. **(Best Presentation)**

[5] Assessing automatic summarization model as a reading assistant
Aujin Kim, Jisu Shin, Soyeong Jeong, Sukmin Cho, and Jong C. Park
Korea Computer Congress (KCC), 2022.

[4] Calibration of Pre-trained Language Model for the Korean Language
Soyeong Jeong, Wonsuk Yang, ChaeHun Park, Jong C. Park
Journal of KIISE (JOK), 2021.

[3] Calibration of Pre-trained Language Model for Korean
Soyeong Jeong, Wonsuk Yang, ChaeHun Park, Jong C. Park
Human & Cognitive Language Technology (HCLT), 2020. **(Best Paper)**

[2] Embedding Academic Majors and Lectures for Analyzing Departments in University
Jinheon Baek, Gwanghoon Jang, Soyeong Jeong, Donghyeon Park, Kiwon Kwon, Jaewoo Kang
Korea Computer Congress (KCC), 2019.

Thesis

[1] Information Retrieval by Augmenting Document Representation
Soyeong Jeong
Master's Thesis, KAIST, 2022

Honors and Awards

Awarded a Ph.D. fellowship from NRF (National Research Foundation) of Korea — 2023
Title: Realistic Open-domain Question Answering System with Large Language Models
Received the Best Paper Award at HCLT 2020 — 2020
Graduated with Honor — 2019
Computer Science and Engineering Department at Korea University
Received the First Prize in the Graduation Project, Competition Among Around 20 Teams — 2019
Computer Science and Engineering Department at Korea University
Nominated as Semester High Honors (Spring 2019) — 2019
Nominated as Semester High Honors (Fall 2018) — 2018
Received the Second Prize for the iOS Hackathon at Korea University — 2018
Awarded as the Finalist, Competition of 21:1 with 735 total submissions — 2017
LG Global Challenger

Academic Services

Reviewer of **ACL ARR 2024** February, April Reviewer — 2024
Reviewer of **ACL ARR 2023** October, December Reviewer — 2023
Reviewer of **TALLIP 2022** (ACM Transactions on Asian and Low-Resource Language Information Processing) — 2022
Reviewer of **ACL-IJCNLP Demo Track 2021** — 2021

Teaching

Computational Linguistics (CS579)
KAIST Fall 2022 (Teaching Assistant)
Natural Language Processing with Python (CS372)
KAIST Spring 2020, Spring 2021, Spring 2023 (Teaching Assistant)

Skills

Languages: Korean (mother tongue), English (fluent)
Programming: Python, C, Java, Swift