

SOYEONG JEONG

◇ Homepage: <https://starsuzi.github.io/> ◇ E-mail: starsuzi@kaist.ac.kr

Last Updated: May 12, 2023

RESEARCH INTERESTS

My research interests are mainly on information retrieval for solving open domain language tasks and interpretation of large language models for making them interpretable when deployed on real-world applications. Not limited to, I am interested in broad topics on natural language understanding.

EDUCATION

Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea

- Ph.D. in School of Computing (Advisor: Jong Cheol Park) *Mar. 2022 - current*
- M.S. in School of Computing (Advisor: Jong Cheol Park) *Mar. 2020 - Feb. 2022*

Korea University, Seoul, Korea

- B.S. in Computer Science and Engineering (Graduated with Honor) *Mar. 2016 - Feb. 2020*
- B.S. in Software Technology and Enterprise Program (Interdisciplinary Program)

Anyang Foreign Language High School, Anyang, Korea

- Major in English *Mar. 2013 - Feb. 2016*

PUBLICATIONS

International Papers

- Phrase Retrieval for Open Domain Conversational Question Answering with Conversational Dependency Modeling via Contrastive Learning
Soyeong Jeong, Jinheon Baek, Sung Ju Hwang, Jong C. Park
Findings of Association for Computational Linguistics (**Findings of ACL**), 2023.
- 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.
- Realistic Conversational Question Answering with Answer Selection based on Calibrated Confidence and Uncertainty Measurement
Soyeong Jeong, Jinheon Baek, Sung Ju Hwang, Jong C. Park
European Chapter of the Association for Computational Linguistics (**EACL**), 2023.
- Augmenting Document Representations for Dense Retrieval with Interpolation and Perturbation
Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang, Jong C. Park
Association for Computational Linguistics (**ACL**), 2022. (**Oral**)
- 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.

- Unsupervised Document Expansion for Information Retrieval with Stochastic Text Generation
Soyeong Jeong, Jinheon Baek, ChaeHun Park, Jong C. Park
Scholarly Document Processing at Conference of the North American Chapter of the Association for Computational Linguistics (**SDP@NAACL**), 2021. (**Oral**)
- 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 Papers

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

- Information Retrieval by Augmenting Document Representation
Soyeong Jeong
Master's Thesis, KAIST, 2022.

RESEARCH EXPERIENCES

NLP*CL (Natural Language Processing and Computational Linguistics) Lab,
Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea
Graduate student (Advisor: Jong Cheol Park) *Mar. 2020 - Current*

- Conducting research on the open conversational question answering task by addressing a problem of an expensive retriever-reader structure with a phrase retrieval model.
- Conducting research on the realistic conversational question answering task by replacing an unrealistic ground-truth history evaluation scheme to confidence- and uncertainty-based answer filtering schemes, which are further calibrated to obtain reliable predictions.
- Conducted research to tackle the problem of the insufficient labeled training data by augmenting dense document representations with interpolation and perturbation on existing dense retrieval methods.
- Conducted research on solving the vocabulary mismatch problem in information retrieval by expanding documents unsupervisedly through diversely generated summaries with pre-trained language models.
- Conducted research on interpreting pre-trained language models based on calibration and analyzed whether the models can capture semantic changes by identifying grammatical characteristics as humans.

DMIS (Data Mining & Information System) Lab, Korea University, Seoul, Korea

Undergraduate Research Intern (Advisor: 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.

ETRI (Electronics and Telecommunications Research Institute), Pangyo, Korea

Research Intern, Seoul SW-SoC Convergence R&BD Center

Jul. 2019 - Aug. 2019

- Participated in the Speech Emotion Recognition project and developed a neural network architecture composed of CNN and LSTM with an Attention mechanism to classify emotions from the input voice.

ETRI (Electronics and Telecommunications Research Institute), Daejeon, Korea

Research Intern, Artificial Intelligence Research Laboratory

Jan. 2019 - Feb. 2019

- Participated in the AIR project (Developing Artificial Social Intelligence for Human-Care Robots) and developed a web crawler for gathering, annotating, and analyzing images and metadata.

HONORS & AWARDS

Best Paper Award

Oct. 2020

32th Annual Conference on Human & Cognitive Language Technology (HCLT)

- Researched pre-trained Korean language models in terms of calibration and smoothing methods.

Semester High Honors

Korea University

- Spring 2019
- Fall 2018

Graduation Project Competition

Jun. 2019

First Prize (Around 20 total submissions), Korea University

- Researched effectiveness of embedding methods on data mining topics such as recommending novel combinations of food ingredients and analyzing university majors with embedding ingredient and curriculum vectors, respectively.

iOS Hackathon

Aug. 2018

Second Prize, Korea University

- Developed a co-buying application based on the iOS platform.

LG Global Challenger

Nov. 2017

Finalist (Competition of 21:1 with 735 total submissions), LG

- Proposed a novel Immersive Virtual Reality (IVR) method to overcome problems of current VR technology (e.g., lacking immersion and vulnerable to motion sickness) using Brain-Computer Interface.

TEACHING

Computational Linguistics (CS579)

Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea

- Fall 2022 (Teaching Assistant)

Natural Language Processing with Python (CS372)

Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea

- Spring 2020, Spring 2021, Spring 2023 (Teaching Assistant)

ACADEMIC SERVICE

Reviewer

- TALLIP 2022
- ACL-IJCNLP Demo Track 2021

SKILLS

Programming Languages : Python, C, Java, JavaScript, Swift
Deep Learning Toolkits : PyTorch, TensorFlow