

Giwon Hong

Technical Research Personnel at KAIST IR&NLP Lab

✉ g.hong@sms.ed.ac.uk  [Google scholar](#)  [Homepage](#)  [LinkedIn](#)

RESEARCH INTERESTS

Natural Language Processing

Question Answering, Data Efficiency, Expert Systems, Language Modeling, Question Generation, Graph QA, Information Retrieval, Interpretability & Explainability

EDUCATION

The University of Edinburgh

PhD student in ILCC program

- Supervisor: Pasquale Minervini (Principal), Edoardo Ponti

Edinburgh, UK
Sep. 2023 - Present

Korea Advanced Institute of Science and Technology (KAIST)

M.S. in School of Computing

- Thesis committee: Sung-Hyong Myaeng, Alice Oh, Meeyoung Cha
- GPA: 3.98 / 4.30 (96.44%)

Daejeon, Korea
Feb. 2018 - Feb. 2020

Sungkyunkwan University (SKKU)

B.S. in Computer Science and Engineering

- GPA: 4.00 / 4.50 (94.3%)
- Major GPA : 4.31 / 4.5 (97.72%)

Suwon, Korea
Mar. 2014 - Feb. 2018

PUBLICATIONS

* indicates equal contribution.

- [1] **Graph-Induced Transformers for Efficient Multi-Hop Question Answering** EMNLP, 2022
Giwon Hong, Jeonghwan Kim, Junmo Kang, Sung-Hyon Myaeng [To appear]
- [2] **Exploiting Numerical-Contextual Knowledge to Improve Numerical Reasoning in Question Answering** Findings of NAACL, 2022
Jeonghwan Kim, Kyung-min Kim, Junmo Kang, **Giwon Hong**, Sung-Hyon Myaeng [pdf]
- [3] **Have You Seen That Number? Investigating Extrapolation in Question Answering Models** EMNLP, 2021
Jeonghwan Kim, **Giwon Hong**, Kyung-min Kim, Junmo Kang, Sung-Hyon Myaeng [pdf]
- [4] **Ultra-High Dimensional Sparse Representations with Binarization for Efficient Text Retrieval** EMNLP, 2021
Kyoung-Rok Jang, Junmo Kang, **Giwon Hong**, Sung-Hyon Myaeng, Joohee Park, Taewon Yoon, Heecheol Seo [pdf]
- [5] **Handling Anomalies of Synthetic Questions in Unsupervised Question Answering** COLING, 2020
Giwon Hong*, Junmo Kang*, Doyeon Lim*, Sung-Hyon Myaeng [pdf]
- [6] **Regularization of Distinct Strategies for Unsupervised Question Generation** Findings of EMNLP, 2020
Junmo Kang*, **Giwon Hong***, Haritz Puerto San Roman*, Sung-Hyon Myaeng [pdf]

- [7] Book chapter **"Finding Datasets in Publications: The KAIST Approach"** Sage London, 2020
In Rich Search and Discovery for Research Datasets
Haritz Puerto-San-Roman, **Giwon Hong**, Minh-Son Cao, Sung-Hyon Myaeng [[Link](#)]
- [8] **Aligning Open IE Relations and KB Relations using a Siamese Network Based on Word Embedding** IWCS, 2019
Rifki Afina Putri, **Giwon Hong**, Sung-Hyon Myaeng [[pdf](#)]

EXPERIENCES

KAIST IR&NLP Lab

July 2020 - July 2023

Technical Research Personnel

- Alternative to mandatory military service (~2023.07.08).
- Working on Question Answering (with Data scarcity, Numbers, and Graphs), Neural IR.
- Person in charge of the Exobrain project, detailed task 1 (KAIST).

KAIST IR&NLP Lab

Mar. 2020 - June 2020

Research Associate

Samsung SDS Senior Data Scientist Course

Feb. 2020 - June 2020

Teaching Assistant

- Class for data processing, analysis, and machine learning (ML) related applications.
- Advising course projects about data analysis and ML techniques.

Korea Advanced Institute of Science and Technology (KAIST)

Mar. 2019 - Dec. 2019

Teaching Assistant

- Teaching assistant for the Text Mining course from probabilistic (e.g., CRF, LDA) to neural-based (e.g., CNN, RNN, LSTM) approaches (2019 1st semester)
- Teaching assistant for the Information Retrieval course (e.g., BM25, PRF, L2R) (2019 2nd semester)

PROJECTS

Development of AI Technology to Support Expert Decision-making that can Explain the Reasons/Grounds for Judgement Results Based on Expert Knowledge

Apr. 2022 - July 2023

Funded by Korean Government (Ministry of Science and ICT)

Hosted by Electronics and Telecommunications Research Institute (ETRI)

- Working on a neuro-symbolic (semi-parametric, KB-based) dynamic learning technology that can effectively model an environment in which knowledge continuously changes.

Exobrain [\[Link\]](#)

Mar. 2018 - Mar. 2023

Funded by Korean Government (Ministry of Science and ICT)

Hosted by Electronics and Telecommunications Research Institute (ETRI)

- The purpose of the research is to provide an **expert-level question answering** service in an environment of the knowledge industry such as law, patents, etc.
- **Participant of Detailed task 3** (2018.03-2019.06)
- **Project manager of Detailed task 3** (2019.06-2019.12)
- **Project manager of Detailed task 1 (KAIST)** (2020.01-Present)
- Researched on extracting KB relations constituting triples for a graph-based QA model [8].
- Lead researcher for an ensemble model that combines the graph-based QA model and reading comprehension QA model (1st rank in the leaderboard of TriviaQA Wikipedia at the date of 08/10/19).
- Worked on solving the anomalies of synthetic questions through inverse BLEU-based paraphrasing and confidence score-based filtering [5].
- Presented a sample-efficient and robust number representation in extrapolation for numerical question answering [2, 3].
- Suggested a method for injecting structural information into the Transformer architecture[1].

Deep Matching for Efficient Search

Mar. 2020 - June 2020

Funded by NAVER Corp.

- **Participant**
- Proposed a novel, efficient and explainable passage retrieval system based on binarized sparse representations that can utilize an inverted index and symbolic techniques [4].

Funded by Korean Government (the Ministry of Science and ICT)

- **Participant**
- Proposed a framework to improve unsupervised question answering by combining different strategies of question generation[6].

HONORS & AWARDS

Rich Context Competition

Feb. 15, 2019

Honorable mention (2nd Place)

- By the Coleridge Initiative at New York University.
- The Rich Context Competition was run by the Coleridge Initiative at New York University and aimed to extract dataset mentions from science publications.
- Finalist (Top 4) in phase 1
- 2nd place in phase 2 (\$2,000)
- Proposed a system to retrieve datasets from papers based on a RCQA model and a question generation. [7].

Scholarship (SKKU)

2014 - 2018

- Jang Young-sil Scholarship (2014 - 2017)
- Academic excellence A (2017 - 2018)

SKILLS

Programming Languages

- Python, C/C++, Java, Javascript

Frameworks & Tools

- PyTorch, PyTorch Lightning, Huggingface, Docker, Codalab, Tensorflow, DGL (Deep Graph Library), NLP Toolkit (SpaCy, NLTK), KBs (Freebase, Wikidata)

English

- TOEFL (iBT): Total 107| Reading 30| Listening 30| Speaking 23| Writing 24

SERVICES

Review Committee - EMNLP 2022, ACL 2023, EMNLP 2023