Junmo Kang

Ph.D. Student at Georgia Tech

✉ junmo.kang@gatech.edu | 🌐 jm-kang.github.io

RESEARCH INTERESTS

Natural Language Processing (NLP)

Large Language Models, Retrieval-Based NLP, Efficiency, Robustness

The long-term goal of my research is to enhance the *practicality* of NLP systems (e.g., Large Language Models) so that they can be widely used *in real-world scenarios*. In particular, my research focuses on *i) how to make NLP models cheaper (i.e., efficient) in terms of data, parameters, or compute*, and *ii) how to build NLP models that are robust to unseen cases in the wild*.

EDUCATION

Georgia Institute of Technology
Ph.D. in Computer Science

Atlanta, GA, USA
Aug. 2022 - Present

- Research Assistant at NLP Lab (Advisors: Dr. Alan Ritter, Dr. Wei Xu)
- GPA: 4.0 / 4.0

Korea Advanced Institute of Science and Technology (KAIST)
M.S. in Computer Science

Daejeon, Republic of Korea
Feb. 2019 - Feb. 2021

- Research Assistant at IR&NLP Lab (Advisor: Dr. Sung-Hyon Myaeng)
- Thesis committee: Dr. Sung-Hyong Myaeng, Dr. Hojin Choi, Dr. Alice Oh
- GPA: 4.03 / 4.30

Chungnam National University
B.E. in Computer Science & Engineering

Daejeon, Republic of Korea
Mar. 2012 - Feb. 2019

- Summa Cum Laude
- GPA: 4.30 / 4.50 (Rank: 1/125 in CSE, Major GPA: 4.41)

PUBLICATIONS

* indicates equal contribution

[1] Schema-Driven Information Extraction from Heterogeneous Tables — Preprint
Fan Bai, **Junmo Kang**, Gabriel Stanovsky, Dayne Freitag, Alan Ritter [pdf]

[2] Discern and Answer: Mitigating the Impact of Misinformation in Retrieval-Augmented Models with Discriminators — Preprint
Giwon Hong*, Jeonghwan Kim*, **Junmo Kang***, Sung-Hyon Myaeng, Joyce Jiyoung Whang [pdf]

[3] Distill or Annotate? Cost-Efficient Fine-Tuning of Compact Models — ACL 2023
Junmo Kang, Wei Xu, Alan Ritter [pdf]

[4] Graph-Induced Transformers for Efficient Multi-Hop Question Answering — EMNLP 2022
Giwon Hong, Jeonghwan Kim, **Junmo Kang**, Sung-Hyon Myaeng [pdf]

[5] Exploiting Numerical-Contextual Knowledge to Improve Numerical Reasoning in Question Answering — Findings of NAACL 2022
Jeonghwan Kim, **Junmo Kang**, Giwon Hong, Kyung-min Kim, Sung-Hyon Myaeng [pdf]

[6] 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]

[7] Leveraging Order-Free Tag Relations for Context-Aware Recommendation
EMNLP 2021

Junmo Kang, Jeonghwan Kim, Suwon Shin, Sung-Hyon Myaeng [pdf]

[8] 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]

[9] Can You Distinguish Truthful from Fake Reviews? User Analysis and Assistance Tool for Fake Review Detection
HCI+NLP@EACL 2021

Jeonghwan Kim*, **Junmo Kang***, Suwon Shin*, Sung-Hyon Myaeng [pdf]

[10] Regularization of Distinct Strategies for Unsupervised Question Generation
Findings of EMNLP 2020

Junmo Kang*, Giwon Hong*, Haritz Puerto San Roman*, Sung-Hyon Myaeng [pdf]

[11] Handling Anomalies of Synthetic Questions in Unsupervised Question Answering
COLING 2020

Giwon Hong*, **Junmo Kang***, Doyeon Lim*, Sung-Hyon Myaeng [pdf]

[12] Let Me Know What to Ask: Interrogative-Word-Aware Question Generation
MRQA@EMNLP 2019

Junmo Kang*, Haritz Puerto San Roman*, Sung-Hyon Myaeng [pdf]

EXPERIENCES

MIT-IBM WATSON AI LAB *Cambridge, MA, USA*
RESEARCH INTERN   May. 2023 - Aug. 2023 (Expected)
- Working on large language models.

GEORGIA TECH NLP LAB *Atlanta, GA, USA*
GRADUATE RESEARCH ASSISTANT   Aug. 2022 - Present
- Worked on zero-shot schema-to-JSON using LLMs [1].
- Worked on cost-efficiency of annotation and distillation [3].

KAIST IR&NLP LAB *Daejeon, Republic of Korea*
RESEARCH ASSOCIATE   Mar. 2021 - Jul. 2022
- Worked on robust retrieval-augmented LMs [2].
- Worked on efficient methods for multi-hop QA [4].

KAIST IR&NLP LAB *Daejeon, Republic of Korea*
GRADUATE RESEARCH ASSISTANT   Feb. 2019 - Feb. 2021
- Worked on question generation and unsupervised question answering for data-efficiency [10,11,12].
- Presented sample-efficient and robust number representations for question answering [8,5].
- Proposed a novel generation model that takes into account the inter-dependency of tags while alleviating the order sensitivity [7].
- Proposed a novel sparse representation model for passage retrieval that can take advantage of an efficient inverted index and symbolic IR techniques [6].

REPUBLIC OF KOREA ARMY *Republic of Korea*
HONORABLY DISCHARGED AS SERGEANT   Apr. 2013 - Jan. 2015
- Compulsory military service.

HONORS & AWARDS

Graduated with Highest Honor in CSE, Chungnam National University	2019
Grand Prize, Business ICT Competition	2018
Excellence Award, Startup Competition	2018
Finalist, NAVER AI Hackathon	2018
Grand Prize, Daejeon Startup School	2017
Best Excellence Award, Startup Picnic	2016
Finalist, Microsoft Imagine Cup Korea	2016