

# SUNKYOUNG KIM

2428, E3-1, 291 Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea 34141

sunkyoung@kaist.ac.kr | wendy.sk.kim@gmail.com | sunkyoung.github.io

Last updated : March 30, 2022

## RESEARCH INTERESTS

---

Large-scale Language Model, Continual Learning, Question Answering, Machine Learning for NLP

## EDUCATION

---

**Korea Advanced Institute of Science and Technology (KAIST)** Mar 2020 - Feb 2022

- M.S. in Computer Science (Advisor : Min-Soo Kim) *Daejeon, Korea*
- Thesis : An Effective Data Selection Method for Pre-training a Korean Language Understanding Model
- Pre-trained language model (InfoLab KorLM) I implemented has ranked 1st in EM score and 4th in F1 score on the KorQuAD v1.0 leaderboard

**Ewha Womans University** Mar 2015 - Feb 2020

- B.S. in Library & Information Science and Multimedia (*Magna Cum Laude*) *Seoul, Korea*
- Minor : Computer Science & Engineering

**Nanyang Technological University** June 2019 - July 2019

- Exchange student (Summer Semester Exchange Programme) *Singapore*

## WORK EXPERIENCE

---

**KAIST InfoLab** Mar 2022 - Present

*Contract Research Scientist* *Daejeon, Korea*

- Develop data selection method for continual pre-training language model
- Develop multi-hop QA system with knowledge representation

## RESEARCH PROJECTS

---

**ITRC Center Student-led Creative Project** Mar 2021 - Dec 2021

- Project title : Efficient text similarity calculation and deduplication study for building neural network of large-scale text graph
- Participated in data processing for deduplication of large-scale News data and evaluation deduplicated text data on pre-training BERT-based model

**KLUE Benchmark Project** Dec 2020 - Jul 2021

- Dataset construction of eight Korean natural language understanding downstream tasks
- Participated in dataset construction of KLUE-RE, KLUE-MRC, modeling of KLUE-MRC, and code refactoring for KLUE-MRC finetuning pipeline

## PUBLICATIONS

---

**KLUE: Korean Language Understanding Evaluation**

- Sungjoon Park, Jihyung Moon, Sungdong Kim, Won Ik Cho, Jiyeon Han, Jangwon Park, Chisung Song, Junseong Kim, Yongsook Song, Taehwan Oh, Joohong Lee, Juhyun Oh, Sungwon Lyu, Younghoon Jeong, Inkwon Lee, Sangwoo Seo, Dongjun Lee, Hyunwoo Kim, Myeonghwa Lee, Seongbo Jang, Seungwon Do, **Sunkyoung Kim**, Kyungtae Lim, Jongwon Lee, Kyumin Park, Jamin Shin, Seonghyun Kim, Lucy Park, Alice Oh, Jung-Woo Ha, and Kyunghyun Cho
- Conference on NeurIPS 2021 Datasets and Benchmarks Track

## **An efficient and effective document deduplication by using similarity based clustering**

- Jungbin Son, **Sunkyoung Kim**, and Min-Soo Kim
- Conference on Korea Computer Congress 2021

## **SCHOLARSHIP**

---

<b>Ewha Future Planning Scholarship</b>	2019
Future Innovation Center, Ewha Womans University	
<b>Study Abroad Program Scholarship</b>	2019
International Affair Office, Ewha Womans University	
<b>Dean's List</b>	2016, 2017, 2019
<b>Samsung Convergence Software Course Excellence Scholarship</b>	2017
Merit-based Scholarship in Ewha-Samsung Convergence Software Course	
<b>Leadership Scholarship</b>	2017
Department of Library and Information Science, Ewha Womans University	
<b>Ewha Volunteering Scholarship</b>	2017
Ewha Womans University	
<b>Yoo So Young Alumni Scholarship</b>	2017
Department of Library and Information Science, Ewha Womans University	

## **TEACHING EXPERIENCE**

---

<b>CS492A Distributed Transactions and Blockchain, KAIST</b>	Fall 2020
Teaching Assistant	

## **TECHNICAL SKILLS**

---

<b>Programming Languages</b>	Python, Shell
<b>Machine Learning Framework</b>	Tensorflow, PyTorch
<b>Tools</b>	Git/GitHub, LaTeX

## **REFERENCE**

---

Prof.Min-Soo Kim, School of Computing, KAIST, minsoo.k@kaist.ac.kr