YUHEUN KIM

ykim72@syr.edu | https://github.com/yuheunk | Syracuse, NY, USA

RESEARCH INTEREST

Natural Language Processing, Artificial Intelligence, Information Extraction, Information Retrieval.

EDUCATION

Ph.D. in Information Science and Technology, Syracuse University *Aug. 2022 – Present*
- Advisor: Bei Yu (Lantis Lab)

M.S. in Digital Analytics, Yonsei University *Mar. 2020 – Feb. 2022*
- Thesis: An Examination of Chronological Datasets Embedding Space for Video-Text Retrieval
- Advisor: Min Song (Deep Text Lab)

B.A. in Economics, Underwood International College (UIC), Yonsei University *Mar. 2015 – Feb. 2020*
- 1 year exchange student at University of California, Irvine *Sep. 2017 – June. 2018*

PUBLICATIONS

1. Hong, G.*, **Kim, Y.***, Choi, Y.*, & Song, M. (2021). BioPREP: deep learning-based predicate classification with SemMedDB. *Journal of Biomedical Informatics*, 122, 103888. [link] [code]

2. Paek, I., Choi, N., Ha, S., **Kim, Y.**, & Song, M. (2022, December). SQ2SV: Sequential Queries to Sequential Videos retrieval. In *2022 IEEE International Conference on Big Data (Big Data)* (pp. 3631-3634). IEEE. [link]

Work in Progress

1. Heo, G.*, Choi, Y., **Kim, Y.**, Hong, G., Song, M., and Yan, E., "A Comprehensive Analysis of Retracted Scientific Publications in the Social Science Domains." (Under Review)

WORK EXPERIENCE

BIGCARE services **Seoul, Korea**

AI Research Intern *Mar. 2022 – Jun. 2022*
- Constructed food nutrition database and built a simple food search algorithm.
- Integrated Django API to food nutrition intake calculating source code.

PROJECT EXPERIENCE

Research Assistant – *Worked in Deep Text Lab under Prof. Min Song (Yonsei University)*

"Biomedical Text Predicate Classification" *Jul. 2020 – Sep. 2021*
- Constructed a biomedical dataset, BioPREP, from SemMedDB which contains biomedical text, filtered entities, and their predicate information.
- Experimented on multiple neural network-based algorithm including CNN, RNN, BERT models and compared, contrasted each model performance in predicting relation from a biomedical text.

- Configured code implementation for fine-tuning pretrained model and uploaded to Github.

"Emerging Issue Detection" *April. 2021 – Sep. 2021*

- Collaboration project with the National Assembly Futures Institute (NAFI).
- Predicted emerging issues for each field of study using term burstiness.
- Analyzed the emerging issue for each field of study and how it was used.

"Keyword Analysis on by-election candidates" *Apr. 2021*

- Collaboration project with a broadcasting network, Channel A.
- Utilized social media data (Naver News, Café comments, Blog posts, YouTube comments, Twitter posts).
- Analyzed social media data for each by-election candidate, using relevant keywords and co-occurrences.
- Conducted a sentiment analysis on each candidate keywords and how the social media data portray their stance on important issues.

"Building AI training dataset for text summarization" *Sep. 2020 – Dec. 2020*

- Constructed AI summarization training dataset in a project hosted by the National Information Society Agency (NIA).

Research Assistant – *Collaborated with Prof. Erjia Yan (Drexel University)*

"Analysis of social science domain retracted papers" *May. 2021 – Oct. 2021*

- Constructed a codebook from retracted papers in social science domain collected from Web of Science.
- Statistically analyzed the pattern for retracted reason.

Project Researcher – *Worked under Prof. Won Suk Lee (Yonsei University)*

"Evaluating user log data for online classes: Preswot" *Feb. 2020 – March. 2020*

- Analyzed student user log data from Preswot, an online class platform.
- Defined three indicators of assessing student performance from the user log data: concentration, participation and comprehension.
- Analyzed the correlation between three indicators and students' academic achievement through a time series analysis.

AWARDS

2020 Digital Analytics Major Competition *Oct. 2020*

1st Place *in Preprocessing competition,* Yonsei University

- Preprocessed datasets using missing value imputation
- Granted scholarship of $1.5K.

2020 Digital Analytics Working Group *Sep. 2020 - Dec.2020*

Awarded scholarship for individual project, Yonsei University

- Trained a set of character images on DCGAN to generate a new character image.
- Granted scholarship of $0.5K.

2020 AI Solution Competition *Oct. 2020 - Nov. 2020*

3rd Place *in Time Series Prediction model*,
Seocho-gu Office & National IT Industry Promotion Agency (NIPA).

TEACHING EXPERIENCE

Teaching Assistant, Syracuse University

- "IST736 (Text Mining)" *Fall 2022*

- Developed sample codes in Topic Modeling as a class material.
- Assisted in grading student assignments and held weekly QA sessions.

Teaching Assistant, Yonsei University

- "Hanhwa Total Employee AI Training" *Spring 2021*

- Instructed a class of 20 Hanhwa Total employees on a text classification task using python.

- "Big Data Youth Campus from Korea Data Agency" *Fall 2020*

- Weekly QA session on students final data analysis project.

Student Mentor, Yonsei University *Summer 2021*

- Instructed 2 students weekly on NLP tasks and algorithmic problems during summer vacation.
- Provided baseline codes for web crawling and text mining tasks such as sentiment analysis.
- Provided instructions and examples for solving coding test problems.

SKILL AND LANGUAGE

Programming: Python, Bash, R, SQL.
Data Science: Pandas, Numpy, Matplotlib.
Deep Learning: Pytorch, Tensorflow+Keras.
Languages: Korean (Native), English (Fluent), Chinese (Beginner), Japanese (Beginner).

[Updated: Feb 2023]