Jaehyeon Myung

Email : mjhbest@kaist.ac.kr
Mobile : +82-10-9980-9381
Github
LinkedIn

Education

- **Korea Advanced Institute of Science and Technologyy** — Republic of Korea, Daejeon
  Master of Science in Computer Science, advised by Meeyoung Cha — *Feb. 2022 – Present*
- **Korea Advanced Institute of Science and Technology** — Republic of Korea, Daejeon
  Bachelor of Science in Computer Science — *Mar. 2017 – Feb. 2022*

Experience

- **Institute for Basic Science(IBS) Data Science Group** — Republic of Korea, Daejeon
  Data Science Researcher — *Jul. 2020 – Present*
  - **Master Student**: Feb. 2021 - Present
  - **Data disparity of COVID-19 in Wikipedia**:
    - Overall analysis of the data disparity in the 11 languages Wikipedia related to COVID-19. This research is published as workshop paper of Companion Proceeding of the Web Conference.
    - Need of information about COVID-19 was increased with a run at the beginning of 2020, but the characteristics of each language version Wikipedia shows differences such as information response speed, size and the contents.
    - From the 11 version of Wikipedia dump file and SQL Query, data such as revision history, length of article, page creation was collected from 2020.
    - Discuss the methodology how to select directly related Wikipedia page to specific subject with embedding and collecting algorithms.
- **Wikimedia Research** — Remote
  Data Science Researcher & Collaborator — *Mar. 2021 – Jan. 2023*
  - **Longitudinal Assessment of Reference Quality on Wikipedia**:
    - Suggest and design the two metrics for evaluating the reference quality in Wikipedia. With the tons of Wikipedia articles, Citation Detective NLP Tool was used for calculating "Citation Need" metric, and reliability of reference source used for "Citation Risk" metric.
    - With Wikipedia's Hadoop system and PySpark, large-scale dataset are collected and databaseized vi SQL Query.
    - Through two indicators that can measure reliability of citations, longitude analysis and user characteristics that affect each citation quality were analyzed. This research was published in the ACM Web Conference 2023.
- **Naver** — Republic of Korea, Seoul
  MLOPs Developer — *Dec. 2021 – Feb. 2022*
  - **MLOPs Pipeline**: Automate the Naver comment filtering model "Clean Bot" to be trained with new data and release. : Python, Git
  - **MLOPs UI Develop**: Develop the UI website for managing trainable data, model version control and performance check. : React, Docker
- **KDI School of Public Policy and Management** — Republic of Korea, Sejong
  Data Scientist/Research Assistant — *Oct. 2021 – Dec. 2021*
  - **Text Analysis of Twitter/Meida Data**: Data & Text analysis (Keyword Network, LDA, Word Embedding) related to 'minimum wage'
  - **Data Processing**: Data crawling and pre-processing from news/social media related to 'minimum wage'
- **Planetarium** — Republic of Korea, Seoul
  BlockChain Engine Engineer — *Dec. 2019 – Feb. 2020*
  - **Apply VRF to Blockchain Server**:
    - VRF(Verified Random Function) technique adaption for Blockchain Platform "Libplanet" and Technical review of performance & Proof of Concept.
    - Implementation test blockchain server for PoC model application : Python/C#.

Publications

- **Longitudinal Assessment of Reference Quality on Wikipedia**: **J. Myung ***, A. Baigutanova*, , A.-J. Chou, D. Saez-Trumper, M. Redi, C. Jung, M. Cha. Proceedings of the ACM Web Conference 2023 (WWW '23) (Published)
- **Algorithmic Assessment of Reference Quality on Wikipedia**: A. Baigutanova, **J. Myung**, A.-J. Chou, D. Saez-Trumper, M. Redi, M. Cha, C. Jung. Proceedings of the International AAAI Conference on Web and Social Media 2022 (ICWSM '22) - Social Science and Sociophysics Track (Accepted)
- **Information flow on COVID-19 over Wikipedia: A case study of 11 languages**: C. Jung, I. Hong, D. Saez-Trumper, D. Lee, **J. Myung**, D. Kim, J. Yun, W. Jung, M. Cha. Companion Proceedings of the Web Conference 2021 (Published)

Technical and Language Skills

- **Programming Languages**: : Python, R, Java, JS, SQL, C/C++, Scala
- **Technologies**: : React, Docker, Pytorch, Linux
- **Languages**: : Korean(Native), English(Fluent), German(Intermediate)