Donghyeon Kim

Senior Research Engineer at Hyundai Motor Company AI Research Lab

donghyeonk@me.com
https://donghyeonk.github.io

RESEARCH INTERESTS

Natural Language Processing, Text Mining, Machine Learning, Bioinformatics

EDUCATION

Korea University, Seoul　　　　　　　　　　　　　　　　　　　March 2011 - Feb 2020
Ph.D. in Computer Science and Engineering　　　　　(On leave: startups, Oct 2014 - Mar 2018)
Advisor: Prof. Jaewoo Kang

Korea University, Seoul　　　　　　　　　　　　　　　　　　　Mar 2003 - Feb 2011
B.S. in Computer Science Education　　　　　(On leave: military service, Aug 2006 - Jul 2008)
Advisor: Prof. Seongbin Park

EXPERIENCE

Hyundai Motor Company　　　　　　　　　　　　　　　　　　　Dec 2019 - Present
Senior Research Engineer　　　　　　　　　　　　　　　　　　　Seoul, Korea

· Natural Language Processing

Konolabs, Inc.　　　　　Apr 2016 - Mar 2018 (Dispatch from Korea Univ. Apr 2016 - Dec 2016)
Data Scientist & Software Engineer　　　　　　　　　　　　　　Seoul, Korea

· Event start time prediction by learning online calendar event titles, week contexts and user preferences
· Scheduling related sentence classification
· Email bot architecture design & implementation

Opinion8, Inc.　　　　　　　　　　　Oct 2014 - Apr 2016 (Dispatch from Korea Univ.)
CTO & Co-founder　　　　　　　　　　　　　　　　　　　　　Seoul, Korea

· Opinion, location, and time based restaurant/attraction/hotel search engine implementation
· Sentiment analysis using dependency trees and part-of-speech tags
· Calculating data, error statistics using MapReduce

Republic of Korea Marine Corps　　　　　　　　　　　　　　　Aug 2006 - Jul 2008
Computer, Signal Corpsman in the 1st division HQ　　　　　　　　Pohang, Korea

· Military service

PUBLICATIONS

Pre-trained Language Models for Biomedical Question Answering
Wonjin Yoon, Jinhyuk Lee, **Donghyeon Kim**, Minbyul Jeong and Jaewoo Kang
BioASQ Workshop at ECML PKDD, Sep 2019

BioBERT: a pre-trained biomedical language representation model for biomedical text mining
Jinhyuk Lee[†], Wonjin Yoon[†], Sungdong Kim, **Donghyeon Kim**, Sunkyu Kim, Chan Ho So and Jaewoo Kang
Bioinformatics, Sep 2019

A Neural Named Entity Recognition and Multi-Type Normalization Tool for Biomedical Text Mining

Donghyeon Kim, Jinhyuk Lee, Chan Ho So, Hwisang Jeon, Minbyul Jeong, Yonghwa Choi, Wonjin Yoon, Mujeen Sung and Jaewoo Kang

IEEE Access, Jun 2019

Learning User Preferences and Understanding Calendar Contexts for Event Scheduling

Donghyeon Kim[†], Jinhyuk Lee[†], Donghee Choi, Jaehoon Choi and Jaewoo Kang

CIKM (full, oral), Oct 2018

BEST: Next-Generation Biomedical Entity Search Tool for Knowledge Discovery from Biomedical Literature

Sunwon Lee[†], **Donghyeon Kim**[†], Kyubum Lee, Jaehoon Choi, Seongsoon Kim, Minji Jeon, Sangrak Lim, Donghee Choi, Sunkyu Kim, Aik-Choon Tan and Jaewoo Kang

PLoS One, Oct 2016

Smith Search: Opinion-Based Restaurant Search Engine

Jaehoon Choi, **Donghyeon Kim**, Donghee Choi, Sangrak Lim, Seongsoon Kim, Youngjae Choi and Jaewoo Kang

WWW (demo), May 2015

Consento: A New Framework for Opinion Based Entity Search and Summarization

Jaehoon Choi, **Donghyeon Kim**, Seongsoon Kim, Junkyu Lee, Sunwon Lee and Jaewoo Kang

CIKM (short), Oct 2012

BOSS: context-enhanced search for biomedical objects

Jaehoon Choi, **Donghyeon Kim**, Seongsoon Kim, Sunwon Lee, Kyubum Lee and Jaewoo Kang

BMC Medical Informatics and Decision Making, Apr 2012

Consento: A Consensus Search Engine for Answering Subjective Queries

Jaehoon Choi, **Donghyeon Kim**, Seongsoon Kim, Junkyu Lee, Sangrak Im, Sunwon Lee and Jaewoo Kang

WWW (poster), Apr 2012

BOSS: a biomedical object search system

Jaehoon Choi, **Donghyeon Kim**, Seongsoon Kim, Sunwon Lee, Kyubum Lee and Jaewoo Kang

DTMBIO (long), Oct 2011

[†]Equal contribution

DOMESTIC PUBLICATIONS

언어 모델링에서의 단어의 효율적인 코드북 표현 (Efficient codebook expressions of word representations in language modeling)

Byungjoo Kim, Yookyung Koh, Chan Ho So, **Donghyeon Kim** and Jaewoo Kang

Korea Computer Congress, Jun 2018

의사결정 지원을 위한 소셜네트워크 서비스의 질의응답 데이터 활용 방안 (Utilizing Social Q&A Data for Decision Making)

Jukyoung Lee, Jaehoon Choi, **Donghyeon Kim**, Donghee Choi, Minhwan Yoo and Jaewoo Kang

The 42nd Annual Meeting and Winter Conference of KIISE, Dec 2015

소셜 빅데이터를 활용한 의사 결정 검색 기술 (Decision making search technology using social big data)

Jaehoon Choi, Donghee Choi, **Donghyeon Kim**, Seongsoon Kim, Sangrak Lim and Jaewoo Kang

Communications of KIISE, Jan 2014

AWARDS

Winner of test batch 1-5, biomedical semantic question answering (Task 7b, Phase B), BioASQ Challenge

Wonjin Yoon, Jinhyuk Lee, **Donghyeon Kim**, Minbyul Jeong and Jaewoo Kang

Sep 2019

The grand prize, Start-up competition, Korea University
Jaehoon Choi, **Donghyeon Kim** and Donghee Choi
Jan 2015

PATENTS

Method, system and non-transitory computer-readable recording medium for providing predictions on calendar
Jaewoo Kang, **Donghyeon Kim**, Jinhyuk Lee and Jaehoon Choi
Application (pending), Korea: 10-2018-0070912, United States: US 2018 / 0174108 A1, Jun 2018

Method, system and non-transitory computer-readable recording medium for providing schedule-related advertisements
Yun Jeong Min, Jaehoon Choi, **Donghyeon Kim** and Jinhyuk Lee
Application (pending), Korea: 10-2018-0070921, United States: US 2018 / 0174191 A1, Jun 2018

SERVICE

WWW (2014, 2013), **CIKM** (2011), **ICWSM** (2014), **ICDE** (2014), **ICDM** (2019), **SIAM** (2011), **BIBM** (2018, 2012, 2011)
Secondary reviewer

Stock study, DMIS Lab., Korea University   Sep 2018 - Jul 2019
Member

- Hybrid Attention Networks (Hu *et al.*, 2018)   `https://github.com/donghyeonk/han`

Paper re-implementation study, DMIS Lab., Korea University   Jul 2018 - Aug 2018
Study group leader

- Match-LSTM (Wang *et al.*, 2016)   `https://github.com/donghyeonk/match-lstm`

- fastText (Joulin *et al.*, 2016)   `https://github.com/donghyeonk/fastText1607`

Data Structure (CNCE210), Korea University   Sep 2011 - Feb 2012
Teaching assistant

SKILLS

Programming Language	Python 3, Java 8+
Natural Language Processing	spaCy, nltk, gensim, Stanford CoreNLP, Google Natural Language, FANSE parser, SentiWordNet, HeidelTime, SentencePiece
Information Retrieval	Apache Solr, Apache Lucene
Machine Learning	PyTorch, TensorFlow, Keras, scikit-learn
Dimension Reduction	t-SNE, UMAP
Data Analysis & Scientific Computing	pandas, NumPy, SciPy
MapReduce	Apache Spark, Apache Hadoop
Database	Amazon DynamoDB, MongoDB, MySQL
Cloud Computing	Amazon Web Services, Google Cloud Platform, IBM Cloud
Project Tracking & Deployment	JIRA, Trello, Jenkins, mixpanel
Visualization	tensorboard, matplotlib, Google Charts, D3.js

PROJECTS

컨센서스 질의응답이 가능한 인공지능비서 시스템 개발   Mar '18 - Dec '19
한국연구재단

인공지능 기반 기업형 일정관리 서비스 상용화   Nov '17 - Mar '18
중소벤처기업부

소셜 빅데이터 기반 콘센서스검색 기술 개발   Dec '14 - Dec '16
연구성과실용화진흥원

문헌 및 실험 빅데이터 통합분석 및 탐색 시스템 개발   Nov '14 - Oct '15
한국연구재단

소셜빅데이터분석과 검색기술의 융합을 통한 차세대 컨센서스검색 플랫폼 개발   Nov '14 - Oct '15
한국연구재단

문서 분할 색인 및 콘센서스랭크 모델을 적용한 소셜미디어 의견기반 객체검색 시스템 개발 May '12 - Apr '15
한국연구재단

학술분야 객체검색시스템 개발   Oct '11 - Jun '12
정보통신기술진흥센터

모바일 사진 컨텐츠 분석을 통해 지능형 멀티미디어 SNS 기능을 극대화 하는 Insight Search Engine의 개발
Jun '11 - May '12
중소기업청

대용량 소셜 미디어의 sentiment 분석을 통한 인사이트 발굴 기술 개발   Mar '11 - Apr '12
한국연구재단

CERTIFICATES

정보처리기사   May '10
한국산업인력공단

컴퓨터활용능력 1급   Feb '10
대한상공회의소

사무자동화산업기사   Aug '09
한국산업인력공단

REFERENCE

Professor Jaewoo Kang
College of Informatics, Korea University
kangj@korea.ac.kr