

# Donghyeon Kim

Ph.D. Candidate in Computer Science and Engineering  
Data Mining & Information Systems Laboratory  
Korea University  
<https://donghyeonk.github.io>

#501B, Woojung Hall of Informatics  
145 Anam-ro, Seoul, Korea 02841  
[donghyeon@korea.ac.kr](mailto:donghyeon@korea.ac.kr)  
<https://github.com/donghyeonk>

## RESEARCH INTERESTS

---

Natural language processing, machine learning, information retrieval, bioinformatics

## EDUCATION

---

### Korea University, Seoul

Ph.D. Candidate in Computer Science and Engineering  
Advisor: Prof. Jaewoo Kang

March 2011 - Feb 2020 (Expected)  
(On leave: startups, Oct 2014 - Mar 2018)

### Korea University, Seoul

B.S. in Computer Science Education  
Advisor: Prof. Seongbin Park

Mar 2003 - Feb 2011  
(On leave: military service, Aug 2006 - Jul 2008)

## EXPERIENCE

---

### Konolabs, Inc.

Data Scientist & Software Engineer

Apr 2016 - Mar 2018 (Dispatch from Korea Univ. Apr 2016 - Dec 2016)  
Seoul, Korea

- Event start time prediction by learning online calendar contexts and user preferences
- Schedule relevant sentence classification using pretrained word embeddings
- Email bot implementation & architecture design

### Opinion8, Inc.

CTO & Co-founder

Oct 2014 - Apr 2016 (Dispatch from Korea Univ.)  
Seoul, Korea

- Entity search service based on opinion, location, and time for travelers
- Sentiment analysis using dependency trees and part-of-speech tags
- Calculating data statistics using MapReduce

### Republic of Korea Marine Corps

Computer, Signal Corpsman in the 1st division HQ

Aug 2006 - Jul 2008  
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, ECML-PKDD, Sep 2019 (in press)

### 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

### BioBERT: a pre-trained biomedical language representation model for biomedical text mining

Jinhyuk Lee<sup>†</sup>, Wonjin Yoon<sup>†</sup>, Sungdong Kim, **Donghyeon Kim**, Sunkyu Kim, Chan Ho So and Jaewoo Kang  
arXiv preprint (arXiv:1901.08746), Jan 2019 (under review)

### **Learning User Preferences and Understanding Calendar Contexts for Event Scheduling**

**Donghyeon Kim**<sup>†</sup>, Jinhyuk Lee<sup>†</sup>, Donghee Choi, Jaehoon Choi and Jaewoo Kang  
CIKM (long, oral), Oct 2018

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

Sunwon Lee<sup>†</sup>, **Donghyeon Kim**<sup>†</sup>, 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

<sup>†</sup>Equal contribution

## **DOMESTIC PUBLICATIONS**

---

### **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

### **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

### **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, BioASQ on biomedical semantic question answering (Task b, Phase B), BioASQ Challenge**

Wonjin Yoon, Jinhyuk Lee, **Donghyeon Kim**, Minbyul Jeong and Jaewoo Kang  
Mar 2019 - May 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), KR20180070912A, US20180174108A1, 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), KR20180070921A, US20180174191A1, Jun 2018

**SERVICE**

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

Secondary reviewer

**Paper re-implementation study, DMIS, 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**

<b>Programming Language</b>	Python 3, Java 8+
<b>Natural Language Processing</b>	spaCy, nltk, gensim, Stanford CoreNLP, Google Cloud Natural Language API, FANSE parser, SentiWordNet, HeidelTime, SentencePiece
<b>Information Retrieval</b>	Apache Solr, Apache Lucene
<b>Machine Learning</b>	TensorFlow, PyTorch, scikit-learn
<b>MapReduce</b>	Apache Spark, Apache Hadoop
<b>Scientific Computing</b>	NumPy, SciPy
<b>Database</b>	Amazon DynamoDB, MongoDB, MySQL
<b>Cloud Computing</b>	Amazon Web Services, Google Cloud Platform, IBM Cloud
<b>Project Tracking &amp; Deployment</b>	JIRA, Trello, Jenkins
<b>Visualization</b>	tensorboard, matplotlib, t-SNE, UMAP

**DOMESTIC RESEARCH PROJECTS**

컨센서스 질의응답이 가능한 인공지능비서 시스템 개발 한국연구재단	Mar '18 - Present
소셜 빅데이터 기반 컨센서스검색 기술 개발 연구성과실용화진흥원	Dec '14 - Dec '16
문헌 및 실험 빅데이터 통합분석 및 탐색 시스템 개발 한국연구재단	Nov '14 - Oct '15
소셜빅데이터분석과 검색기술의 융합을 통한 차세대 컨센서스검색 플랫폼 개발 한국연구재단	Nov '14 - Oct '15
문서 분할 색인 및 컨센서스랭크 모델을 적용한 소셜미디어 의견기반 객체검색 시스템 개발 한국연구재단	May '12 - Apr '15
학술분야 객체검색시스템 개발 정보통신기술진흥센터	Oct '11 - Jun '12

모바일 사진 콘텐츠 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

## REFERENCE

---

**Professor Jaewoo Kang**

College of Informatics, Korea University

kangj@korea.ac.kr