

Hongjin Qian

PhD. student at
Gaoling School of Artificial Intelligence
Renmin University of China

ian@ruc.edu.cn
chien.io
+86 13051011770

Summary

I am a third-year PhD student from GSAI, RUC, working on NLP and IR. My research interests focus on conversational search, retrieval-augmented generation, and user modelling. I have several publications in top-tier conferences (e.g. SIGIR, CIKM and EMNLP). Before enrolling into GSAI, I was an NLP engineer of Beijing Academy of Artificial Intelligence (BAAI), working on QA system. Prior to BAAI, I worked in a start-up AI company, Elensdata.

Education

Renmin University of China

PhD. Student in Artificial Intelligence (Supervised by Prof. Zhicheng Dou and Prof. Ji-Rong Wen)

Beijing, China

Sept. 2020 – Present

University of Sydney

Master of Information Technology, Major in Data Management and Analytics

Sydney, Australia

Jul. 2017 – Jan. 2019

Nankai University

Bachelor of Science, Major in Electronic information Science and Technology

Tianjin, China

Sept. 2013 – Jun. 2017

Experience

Research Intern

Poisson Lab, Huawei

- WebBrain
- Model-oriented IR

Apr. 2022 – Present

Beijing, China

PhD. Researcher

Gaoling School of Artificial Intelligence, Renmin University of China

- Personalized Chatbot
- Conversational Information seeking
- Information Retrieval

Sept. 2020 – Present

Beijing, China

NLP Engineer

Beijing Academy of Artificial Intelligence Mentor: Zhanliang Liu

- QA System dedicated in Governance Domain
- Dense Retrieval
- Fine-Grained Named Entity Recognition

Jun. 2020 – Mar. 2021

Beijing, China

NLP Engineer

Elens.AI Mentor: Zhanliang Liu

- Core NLP Toolkit for Chinese, English and Uyghur
- NLP Applications in Multiple Domains (Financial, Security, Media etc.)
- Large-Scale Pretrained Language Model and Text Generation

Jan. 2019 – Sept. 2020

Beijing, China

Publications

* indicates equal contribution.

- (1) **Learning Denoised and Interpretable Session Representation for Conversational Search. theWebConf 2023**
Kelong Mao, **Hongjin Qian**, Fengran Mo, Zhicheng Dou, Bang Liu, Xiaohua Cheng, Zhao Cao
- (2) **Topic-Enhanced Personalized Retrieval-based Chatbot. ECIR 2023**
Hongjin Qian and Zhicheng Dou
- (3) **Explicit Query Rewriting for Conversational Dense Retrieval. EMNLP 2022**
Hongjin Qian and Zhicheng Dou
- (4) **ConvTrans: Transforming Web Search Sessions for Conversational Dense Retrieval. EMNLP 2022**
Kelong Mao, Zhicheng Dou, **Hongjin Qian**, Fengran Mo, Xiaohua Cheng and Zhao Cao
- (5) **Less is More: Learning to Refine Dialogue History for Personalized Dialogue Generation. NAACL 2022**
Hanxun Zhong, Zhicheng Dou, Yutao Zhu, **Hongjin Qian** and Ji-Rong Wen

- (6) **Webformer: Pre-training with Web Pages for Information Retrieval. SIGIR 2022**
Yu Guo, Zhengyi Ma, Jiaxin Mao, **Hongjin Qian**, Xinyu Zhang, Hao Jiang, Zhao Cao and Zhicheng Dou
- (7) **Curriculum Contrastive Context Denoising for Few-shot Conversational Dense Retrieval. SIGIR 2022**
Kelong Mao, Zhicheng Dou and **Hongjin Qian**.
- (8) **Learning Implicit User Profile for Personalized Retrieval-Based Chatbot. CIKM 2021**
Hongjin Qian, Zhicheng Dou, Yutao Zhu, Yueyuan Ma, and Ji-Rong Wen
- (9) **Pchatbot: A Large-Scale Dataset for Personalized Chatbot. SIGIR 2021**
Hongjin Qian, Xiaohe Li, Hanxun Zhong, Yu Guo, Yueyuan Ma, Yutao Zhu, Zhanliang Liu, Zhicheng Dou and Ji-Rong Wen
- (10) **Speaker or Listener? The Role of a Dialog Agent. Findings of EMNLP 2020**
Yafei Liu*, **Hongjin Qian***, Hengpeng Xu, Jinmao Wei

Granted Patents

- (1) **Latin letter-based Uyghur language processing method and system (CN111428509A)**
Hongjin Qian, Zhen Huang, Zhicheng Dou, Zhanliang Liu
- (2) **Method and electronic equipment for searching long text in question-answering task in open field (CN111881264A)**
Hongjin Qian, Zhanliang Liu, Zhicheng Dou, Jiajun Liu
- (3) **Optical character recognition method and device fused with language model and electronic equipment (CN111738251A)**
Hongjin Qian, Zhanliang Liu, Zhicheng Dou, Jiajun Liu
- (4) **Uyghur language entity identification method and device and electronic equipment (CN111814433A)**
Hongjin Qian, Zhanliang Liu, Zhicheng Dou, Jiajun Liu
- (5) **Multi-level-length text vector retrieval method and device and electronic equipment (CN112988952A)**
Hongjin Qian, Zhanliang Liu, Zhicheng Dou, Gang Cao, Ji-Rong Wen
- (6) **Semantic analysis method and device based on rules and learning and electronic equipment (CN112347793A)**
Hongjin Qian, Xiaotong Li, Zhanliang Liu, Yushu Yang, Zhicheng Dou, Gang Cao, Ji-Rong Wen
- (7) **Index data-based natural language processing method and system (CN111488423A)**
Zhanliang Liu, **Hongjin Qian**, Zhicheng Dou, Jiajun Liu
- (8) **Multi-data-source NL2SQL system based on semantic rules and multi-dimensional model (CN112559550A)**
Zhi Li, **Hongjin Qian**, Zhanliang Liu
- (9) **Government affair FAQ knowledge base automatic construction method and device and electronic equipment (CN112784022A)**
Sixu Guo, **Hongjin Qian**, Yushu Yang, Zhanliang Liu, Zhicheng Dou, Gang Cao, Ji-Rong Wen
- (10) **FAQ knowledge base automatic generation method and device based on complex data types (CN112800177A)**
Sixu Guo, **Hongjin Qian**, Yushu Yang, Zhanliang Liu, Zhicheng Dou, Gang Cao, Ji-Rong Wen
- (11) **Text retrieval result scoring method, retrieval method and device (CN111930928A)**
Yu Zhang, **Hongjin Qian**, Zhanliang Liu, Zhicheng Dou
- (12) **Text extraction method and device for image document and electronic equipment (CN112036406A)**
Yuanyuan Huang, **Hongjin Qian**, Zhanliang Liu, Zhicheng Dou
- (13) **Semantic retrieval method and device and electronic equipment (CN112035730A)**
Yang Zhou, **Hongjin Qian**, Zhanliang Liu, Zhicheng Dou
- (14) **Method and device for rapidly realizing NL2SQL based on vectorization semantic rule (CN112001188B)**
Chaofeng Xiao, Zhi Li, **Hongjin Qian**, Zhanliang Liu, Zhicheng Dou
- (15) **Interface performance testing method and device and electronic equipment (CN111881060A)**
Huan Zhang, Zhi Li, **Hongjin Qian**, Zhanliang Liu
- (16) **FAQ knowledge base automatic construction method and device based on table data (CN112800032A)**
Sixu Guo, Yushu Yang, **Hongjin Qian**, Zhanliang Liu, Zhicheng Dou, Gang Cao, Ji-Rong Wen

Pending Patents

- (1) **A kind of vocabulary sign method and device based on the Chinese phonetic alphabet (CN110162789A)**
Zhicheng Dou, **Hongjin Qian**, Zhen Huang
- (2) **Retrieval type personalized dialogue method and system (CN113901188A)**
Zhicheng Dou, **Hongjin Qian**

Service

PC Reviewer: EMNLP 2021, EMNLP 2022, EACL 2023, ACL 2023

Tutoring: Data Structures, Academic Writing

Awards

Outstanding Innovative Talents Cultivation Funded Programs 2022

[Top 10%] at Renmin University of China

Outstanding Innovative Talents Cultivation Funded Programs 2021

[Top 10%] at Renmin University of China

First-Class Academic Scholarship 2021

[Top 30%] at Renmin University of China

Technical Skills, Language Skills

Programming Languages: Python, C, Go, SQL

Framework: Pytorch, Tensorflow, PaddlePaddle, ElasticSearch, Logstash, Kibana

Platform: AWS, GCP, Azure, Segment.io, Mixpanel, ModelArts

Writing: \LaTeX , Office

Languages: Chinese (native), English (fluent)

Last updated: Apr, 2023