

# 董骞

✉ qiangdong.97@qq.com · ☎ (+86) 185-1032-0939 · 🌐 Qian · 🏠 个人主页

## 👤 个人简介

我的主要研究方向包括无监督大模型对齐、检索增强生成、知识增强模型、文档排序及法律文本检索等。在 SIGIR、WWW 等顶级会议发表多篇论文。

## 🎓 教育背景

清华大学, 北京	2022 – 至今
博士在读 计算机科学与技术	
研究方向: 大语言模型, 偏好对齐, 检索增强生成, 文本排序	
导师: 马少平教授、刘奕群教授、艾清遥副教授	
中国科学院软件研究所, 北京	2019 – 2022
硕士 计算机应用技术	
研究方向: 预训练语言模型, 文档排序	
导师: 牛树梓研究员	
华南理工大学, 广州, 广东	2015 – 2019
学士 软件工程	

## 💼 工作经历

小红书, 实习	2024 年 11 月 – 至今
新加坡国立大学, NExT++ 研究中心, 访问学者	2024 年 8 月 – 2024 年 10 月
清华大学, 计算机系计研 41 团支部, 团支书	2022 年 8 月 – 2024 年 8 月
腾讯, 实习	2022 年 8 月 – 2022 年 10 月
百度, 实习	2021 年 7 月 – 2024 年 10 月

## 📄 研究成果

<b>SimRACG: A Simple Retrieval-Augmented Code Generation Paradigm</b>	2025
<ul style="list-style-type: none"><li>作者: Qian Dong, et al.</li><li>会议: Under Review</li><li>核心贡献: 设计了一种简化的检索增强代码生成范式, 提高了代码生成质量, LLM-based embedder 训练成本降低超 90%。</li></ul>	
<b>Qilin: A Multimodal Information Retrieval Dataset with APP-level User Sessions</b>	2025
<ul style="list-style-type: none"><li>作者: Qian Dong*, Jia Chen*, et al. (共同一作)</li><li>会议: SIGIR 2025(CCF-A)</li><li>核心贡献: 高质量多模态信息检索数据集, 涵盖 APP 级用户会话信息, 支持多种信息检索任务。</li></ul>	
<b>DecoupledRAG: An Efficient and Effective Retrieval Augmented Generation Framework via Cross Attention</b>	2025
<ul style="list-style-type: none"><li>作者: Qian Dong, et al.</li><li>会议: WWW 2025 (CCF-A)</li><li>核心贡献: 提出了一种解耦知识与上下文的方法, 提高了 RAG 系统的效率和有效性, 知识注入的 prefilling 成本从平方降到线性。</li></ul>	

## Unsupervised Large Language Model Alignment for Information Retrieval via Contrastive Feedback 2024

- 作者: Qian Dong, et al.
- 会议: SIGIR 2024 (CCF-A)
- 核心贡献: 提出了基于对比反馈的偏好对齐方法, 提高大语言模型在信息检索任务中的表现。

## $T^2$ Ranking: A Large-scale Chinese Benchmark for Passage Ranking 2023

- 作者: Xiaohui Xie, Qian Dong, et al. (学生一作)
- 会议: SIGIR 2023 (CCF-A)
- 核心贡献: 目前最大的中文段落排序基准数据集, 已成为 BGE、gte-Qwen 等模型训练的基础数据。

## $I^3$ Retriever: Incorporating Implicit Interaction in Pre-trained Language Models for Passage Retrieval 2023

- 作者: Qian Dong, et al.
- 会议: CIKM 2023 (CCF-B)
- 核心贡献: 通过隐式交互, 提高 BERT 的文本检索能力, 在多个基准上达到 SOTA。

## Incorporating Explicit Knowledge in Pre-trained Language Models for Passage Re-ranking 2022

- 作者: Qian Dong, et al.
- 会议: SIGIR 2022 (CCF-A)
- 核心贡献: 结合知识图谱提高 BERT 在段落重排序任务中的表现, 2021 年在 MS-MARCO 上排名全球第二。

## Disentangled Graph Recurrent Network for Document Ranking 2022

- 作者: Qian Dong, et al.
- 会议: Data Science and Engineering-DASFAA2021 Special Issue (JCR-Q1, IF: 7.2)
- 核心贡献: 采用因果推理方法消除文档中的无关信息, 提高文档排序质量。

## Legal Judgment Prediction via Relational Learning 2021

- 作者: Qian Dong, et al.
- 会议: SIGIR 2021 (CCF-A)
- 核心贡献: 采用关系学习方法提高法律判决预测的准确性。

## Latent Graph Recurrent Network for Document Ranking 2021

- 作者: Qian Dong, Shuzi Niu
- 会议: DASFAA 2021 (CCF-B)
- 核心贡献: 设计了隐式图循环网络, 有效建模查询与文档内部结构关系, 提升排序性能。

## ✿ 学术服务

- Reviewer/PC-member: SIGIR, ACL, CIKM, WSDM, WWW, COLING

## ♡ 获奖情况

清华大学综合奖学金	2023 年-2025 年
国家奖学金	2021 年 11 月
国科大学业奖学金	2021 年 9 月
优秀共青团员	2021 年 5 月
校级三好学生	2021 年 5 月