

# 王天舒

✉ [contact@wang.tianshu.me](mailto:contact@wang.tianshu.me) | [github.com/tshu-w](https://github.com/tshu-w) | 📍 北京

在线版本

## 教育经历

- 中国科学院软件研究所 · 国科大杭高院, 计算机科学与技术, 博士 2019年9月 - 2025年7月
- 研究方向: LLM, Entity Resolution, Text2SQL 和 Structured Text Retrieval 导师: 孙乐, 韩先培研究员
  - 平均绩点: 3.73/4, 获得荣誉: 三好学生, 一等学业奖学金 硕转博
- 北京交通大学, 计算机科学与技术, 学士 2015年9月 - 2019年7月
- 平均绩点: 91.2/100 (前 10%), 英语六级: 513, 推荐免试攻读研究生

## 科研经历

- Tianshu Wang *et al.*, “Bridging the Gap between Reality and Ideality of Entity Matching: A Revisiting and Benchmark ReConstrction”, in *Proc. of IJCAI 2022*, pp. 3978–3984. doi: [10.24963/ijcai.2022/552](https://doi.org/10.24963/ijcai.2022/552). CCF-A
- 分析实体匹配基准隐含的不合理假设, 构建首个开放实体、不平衡标签和多模态基准, 以缓解评估偏差
- Tianshu Wang *et al.*, “DBCopilot: Scaling Natural Language Querying to Massive Databases”, *CoRR*, 2023, doi: [10.48550/arXiv.2312.03463](https://doi.org/10.48550/arXiv.2312.03463). [tshu-w/DBCopilot](https://github.com/tshu-w/DBCopilot) ☆ 48 stars 投稿至 ICDE 2025
- 提出一种大小模型协同合作, 生成式联合检索的 Text2SQL 框架, 将自然语言查询扩展到海量数据库
- Tianshu Wang *et al.*, “Match, Compare, or Select? An Investigation of Large Language Models for Entity Matching”, *CoRR*, 2024, doi: [10.48550/arXiv.2405.16884](https://doi.org/10.48550/arXiv.2405.16884). 投稿至 ACL ARR 2024 June
- 探究 LLM 用于实体匹配的不同策略, 并基于探究结果设计了一种更有成效和效率的组合式框架
- Tianshu Wang *et al.*, “Towards Universal Dense Blocking for Entity Resolution”, *CoRR*, 2024, doi: [10.48550/arXiv.2404.14831](https://doi.org/10.48550/arXiv.2404.14831).
- 基于 GitTables 训练通用实体记录 (Record) 表示模型, 以实现无需特定领域数据微调的稠密分块

## 工作经历

- AI Lab 实习生 | 字节跳动 2019年2月 - 2019年8月
- 协助头条搜索「精准问答」项目的开发及优化
- 使用 Spark & Hive 统计每日 Top 10 查询的精准问答覆盖情况, 为正向挖掘提供支持
  - 通过 MapReduce 对全量 CQA 数据根据规则、答案类型、问题意图、答案质量等维度进行筛选
  - 使用动态规划算法优化答案片段抽取模块, 提升 >50% 的答案摘要抽取效率

## 荣誉奖项

- 国际大学生程序设计竞赛 (ICPC) 亚洲区域赛 2016年 - 2018年
- 金奖 × 2: 第 42 届北京站, 第 43 届焦作站
  - 银奖 × 3: 第 42 届乌鲁木齐站, 第 42、43 届东部赛区总决赛
  - 铜奖 × 1: 第 41 届东部赛区总决赛
- 中国大学生程序设计竞赛 (CCPC) 2017年 - 2018年
- 季军 × 1: 第 4 届全国邀请赛湘潭站
  - 金奖 × 3: 第 3 届秦皇岛站, 第 4 届桂林站、吉林站

## 专业技能

- 技术特长: 熟悉 LLM 训练及 Agent 应用, 信息检索, 表示学习, 软件开发, DevOps, 数据分析等
- 编程语言: 掌握 Python, C++, Bash, 熟悉 SQL, Lua, TypeScript, Java
- 系统软件: 熟练使用 macOS/Linux, Emacs, Git, Pytorch, Docker, Ray, Spark, L<sup>A</sup>T<sub>E</sub>X 等
- 开源贡献: 向 [Lightning-AI/pytorch-lightning](https://github.com/lightning-ai/pytorch-lightning) (10 commits) 等 33 个开源项目多次贡献代码

最后更新于 2024 年 09 月 03 日