

杨桂森 | Munger Yang

中国, 北京市 - 海淀区 - 北京市海淀区树村路 19 号中国科学院信息工程研究所, 100085

☎ 15803173971 • ✉ jianghuxiaomiao666@163.com

🌐 <https://www.mungeryang.top> • 🐱 @Mungeryang • 🐼 @迟钝爬爬

教育背景

中国科学院大学/中国科学院信息工程研究所

硕士, 网络空间安全-大数据存储与管理

2024 年 8 月至今

- 研究方向: 大数据存储与智能计算、检索增强生成、实时计算、分布式存储
- 中国科学院信息工程研究所-06 智能认知安全实验室-1G1
- 指导教师: 李斌斌老师

河北大学

学士, 信息管理与信息系统

2020 年 9 月 - 2024 年 6 月

- 核心课程: 数据结构、数据库、计算机网络、操作系统、信息管理学、信息检索、运筹学
- GPA: 4.32/5.0(2/90), 推免至中国科学院大学

荣誉

全国大学生数学竞赛

一等奖

2024 年 3 月

全国大学生数学建模竞赛

河北省一等奖

2023 年 10 月

证书

阿里云认证

云计算与大数据助理工程师

2024 年 5 月

华为认证

华为云存储初级工程师

2024 年 6 月

项目

数据采集

基于 Scrapy 的 boss-X 聘网站信息爬取

https://www.github.com/Mungeryang/boss_data

2023 年 11 月 - 2023 年 12 月

- 任务: 使用 Scrapy 框架采集 boss 直聘网站互联网/AI 领域全站岗位信息
- 项目难点: 分页爬取、selenium 动态页面加载、parse 函数请求传参、mysql 数据库存储
- 技术栈: Scrapy、selenium、Http 协议、XPath 数据解析、异步爬取

离线数仓

在线教育平台离线数仓开发

https://www.github.com/Mungeryang/data_ware

2023 年 6 月 - 2024 年 2 月

- 任务: 采用主流大数据技术和工具搭建知行教育大数据分析平台
- 项目难点: 维度建模分析、Hive SQL 优化、拉链表、分桶表设计
- 技术栈: Hadoop、MapReduce、Hive、Zookeeper、Yarn

和鲸社区- AI 夏令营

大模型应用技术实践

2024 年 7 月 - 2024 年 8 月

- 大模型应用开发实践 & 能力认证
- RAG 原理与应用
- 微调技术原理与应用

大学生创新创业项目实训

面向大数据的农产品外贸风险识别与评估

2022 年 10 月 - 2023 年 6 月

- 负责 Python 爬虫进行数据源采集, 使用 MySQL 数据库进行数据存储, 数据量 6 万条
- 负责使用 KMEANS、PCA 非监督机器学习算法, 挖掘出特征标签, 搭建农产品外贸风险识别模型