

联系方式

- 手机: 18629508095
 - Email: leoluo22@hotmail.com
 - 微信: luohao_Vamp
-

个人信息

- 罗皓/男/1995
 - 本科/长安大学 计算机科学与技术
 - 工作年限: 1年
 - 技术博客: <http://blog.csdn.net/u013271714>
 - Github: <http://github.com/leoluo22>
 - 期望职位: 爬虫工程师
-

工作经历

中国银行软件中心（2017年7月 ~ 至今）

手机银行搜索项目

中国银行手机银行端的搜索，目前支持对已有的功能和产品进行搜索。使用了开源的搜索框架Solr，后台基于Spring+SpringMVC+MyBatis，使用dubbo+zookeeper作为RPC框架以及Maven作为项目管理工具。我在前期负责了搜索框架的搭建（服务器搭建，索引导入，分词策略）以及后台接口的实现（搜索，索引及词库的同步和维护等）。目前负责中国银行手机银行端的搜索，主要工作是对搜索体验进行优化。

- 为了减少首次启动项目时连接到搜索服务器的时间过长，将项目与搜索服务器连接的建立放在了StartupListener，大大减少了等待时间。
- 通过查询官方英文文档，避免了每次更新搜索配置时都需要重启服务器的操作。
- 对外援写的代码进行重构，使得主要逻辑里的代码量减少了一半。

产品趋势项目

跟踪手机银行和E融汇两个产品在各大主流应用商店的评论。在此项目中我独立负责爬虫的设计以及后台接口的开发。爬虫主要使用requests和Beautifulsoup实现，而后台基于Django+MySQL+Redis，提供了RESTful的接口。前端主要展现了指定时间段内APP的评分，正负面评论，以及关键词。

- 使用jieba分词提取评论中的关键字，去除无意义的词，展示数量前十的正负面关键词。
- 为了获得用户评论的情感，一开始我使用了SnowNLP来获取用户情感是正面的概率，但是效果不佳，后来调用了百度提供的情感分析API来实现。
- 一开始后台直接调用的MySQL进行查询，但是速度太慢，于是加入了Redis作为缓存，使得访问速度提高了96%。

指标监控

通过分析每天9台服务器上的日志（日志从2G到5G）监控每天手机银行客户端的打开数，登录数，总请求数，最大并发量，各个接口访问数以及错误数等。

其他

- 获得2017年度中国银行软件中心优秀新人奖
- 在创新技能大赛担任队长，以“影响力指数分析系统”获得“最佳展示奖”

项目和作品

开源项目

- [PyModel](#): 使用Python编写的一个ORM库
- [notesmail](#): 使用Python操作Lotus Notes

爬虫项目

- 拉勾网职位数据采集: 基于Scrapy-Redis实现的分布式爬虫，自动更换代理，使用MongoDB存储。一共抓取了110w+条数据，进行了数据分析和可视化实现。
- 学校图书馆书籍信息采集: 基于requests和beautifulsoup实现。

技术文章

- [使用Python发送和读取Lotus Notes邮件](#)
- [如何在SolrCloud中添加删除节点](#)
- [内核内存泄露，Intel处理器设计缺陷迫使Linux, Windows重新设计\(翻译文章\)](#)

演讲和讲义

- 2017年10月公司内部分享: [基于Solr的企业级搜索](#)

技能清单

- CET6: 529
- 编程语言: Python/Java
- 框架: Django/Scrapy/Spring
- 数据库相关: MySQL/Oracle/SQLite/Redis/MongoDB
- 版本管理、文档和自动化部署工具: Svn/Git
- 单元测试: junit, unittest