
APP评论分析系统项目

—目标与服务模型—

创建日期: 2017-10-27

最后修订日期: 2017-11-29

作者:	项目经理
日期:	2017-10-27
审核:	项目经理
日期:	2017-11-29

目录

1. 项目目标	2
1.1. 项目背景	2
1.1.1. 公司背景	2
1.1.2. 技术背景	2
1.1.3. 业务背景	2
2. 项目价值	3
2.1. 项目难点分析	3
2.1.1. 评论信息的获取	3
2.1.2. 语法分析	3
2.1.3. 提供评论分析结果和建议	3
2.2. 项目优势分析	3
2.2.1. 技术优势	3
2.2.2. 应用优势	4
3. 项目解决思路	4
3.1. 总体思路	4
3.2. 具体做法	4

1. 项目目标

1.1. 项目背景

1.1.1. 公司背景

虹软公司是全球领先的专业计算摄影与计算机视觉技术公司。创建于 1994 年，总部设在美国加利福尼亚硅谷，同时在欧洲、东京、首尔、台北、上海、杭州、南京都设有区域性的商业与研发基地。

多年来，一直专注于计算摄影与计算机视觉技术领域的虹软公司，结合市场需求并引领技术趋势，不断自主研发和创新，拥有此领域强势的核心技术能力，已为全球数十亿台的硬件产品提供了解决方案，给全球消费者带来了更好的用户体验和真正的价值。

虹软的计算摄影与计算机视觉技术广泛应用于以智能手机为主的各种设备和平台，也延伸到智能电视、机器人、智能冰箱、智能汽车、无人机、扫地机、互联网应用等；为三星、LG、索尼、亚马逊、日立、佳能、尼康、奥林巴斯、松下、联想、小米、格力、美的、乐视、Vivo、金立等全球知名的设备制造商的产品提供个性化的计算摄影与计算机视觉解决方案；也为互连网应用提供了差异化和更好的用户体验及技术服务。

虹软公司，80%以上的成员为科学工作者及工程师，成功聚集了比肩世界最有影响力的视觉领域专家，并吸纳和培养了来自国内、国外一流高校的优秀人才作为生力军。虹软公司凭借在技术、人才、专利上近二十年的积累，在大数据、深度学习、AR&VR 及人工智能领域都有深入研究和创新，为全球用户开启新的智慧生活。

1.1.2. 技术背景

德鲁伊的翡翠梦境团队所设计的评论数据分析系统使用了爬虫技术、分词技术、评论分析技术、数据可视化技术等一系列行业新兴技术。通过对海量的评论数据进行语法分析后，通过趋势线，词云图，饼图等相关图表手段形式直观地呈现给用户。使用户直观、准确的了解 APP 的市场状况。

1. 爬虫技术

在现实生活中，我们想把 APPStore 中的内容保存下来通常只能通过截图或者复制粘贴。而用 python 的 requests 库可以非常轻松的实现一个简单的爬虫功能，将海量的评论数据抓取到本地并保存到数据库。

2. 分词技术

3. 数据可视化技术

用户很难的对一堆复杂的，冗长的数据做出合理，准确的分析，即使系统展现给用户的是一系列经过分析的统计数据，如下载量、评论量增长等。而趋势线，词云图，饼图等相关图表手段正好解决了这个问题，用户通过这一系列的图表可以准确、直观的获取 APP 的市场运营状况

1.1.3. 业务背景

随着智能机的普及，APP 的开发和使用层出不穷；同时加之网络的迅猛发展，改变了用户表达自我意见的渠道。对于每一款手机应用，都会在不同的平台做发布，很多用户会在各个平台评论该款的使用感受，这些感受因为来自直接用户的评论，显得十分重要。这些评论信息不仅能让用户了解现在的 APP 市场，而且更重要的是能够为产品生产提供宝贵的反馈信息，精确地获取到对企业有用的信息以便改进产品的

质量。然而，人们不可能逐条地有效阅读海量的用户评论，且每个平台的用户在语言、用词等方面都会有较大的不同，因此，一款基于评论的分析系统显得尤为重要。

近几年，也陆续出现了多款 APP 市场分析系统，如：企鹅风讯、talkingdata 等，他们大多更加偏重于 APP 的排行、下载量等市场运营数据，而对于评论的分析、用户的体验、结果反馈上不是特别看重。而本系统旨在给用户提供反馈信息使得他们能够精确抓取信息并作出相应的调整，因此对数据的深入分析和直观呈现是本系统的重点。

为了更好地迎合企业的需求，为企业提供个性化服务，我们团队致力于 APP 评论数据分析系统的解决和完善。

2. 项目价值

2.1. 项目难点分析

2.1.1. 评论信息的获取

本系统实现的前提条件是获取海量的评论数据，因此爬虫成为该项目第一个难点。由于各个应用商店的页面展现形式不同，因此爬虫系统需要适应各个不同的应用商店。爬取得到的数据格式不同也是爬虫系统面临的一个重大问题，爬虫系统需要对不同应用商店爬取到的数据做一个简单处理，保证插入数据库的数据格式、长度合适。

2.1.2. 语法分析

目前市面上所有支持语法分析的库（比如 R 语言和 python 的库）都无法很好的执行对评论数据的语法分析，如何完成语法分析是我们面临的一大难点。

2.1.3. 提供评论分析结果和建议

本系统实现的最终目标是在评论分析的基础上对用户的 APP 提出改善建议。该难点将是伴随项目整个设计阶段以及后期维护阶段所要解决的最大难点

2.2. 项目优势分析

2.2.1. 技术优势

2.2.1.1. 爬虫技术实现海量数据的抓取

Python 语言对爬虫技术支持经过长时间的发展已经臻至化境，系统完全可以通过爬虫技术获取海量的评论分析数据。

2.2.1.2. 前后端数据交互的精简化

Spring-boot 框架内置 thymeleaf 渲染方式简化了前后端交互的过程，前后端交互更加方便简洁。

2.2.1.3. 持久化框架 mybatis 的应用

Mybatis 优势：

1. 易于上手和掌握。
2. sql 写在 xml 里，便于统一管理和优化。
3. 解除 sql 与程序代码的耦合。
4. 提供映射标签，支持对象与数据库的 orm 字段关系映射
5. 提供对象关系映射标签，支持对象关系组建维护
6. 提供 xml 标签，支持编写动态 sql
7. 支持逆向工程，不需要再写各种 sql 语句

2.2.2.应用优势

2.2.2.1.有针对性的用户权限管理平台

管理员可以对用户进行分组，以确定用户所拥有关注 APP 的权利，此功能是为了保护某些测试阶段未上线 APP。

2.2.2.2.个性化的用户 APP 管理平台

用户可以自行关注 APP，可以对各个 APP 实时的添加相关分词，使得每个用户对每个 APP 都有一个个性化的词库，在搜索评论时，有限显示分词。

2.2.2.3.数据统计分析可视化功能

本系统对爬取的大量数据进行分析处理，将处理结果用形式多样的图表表示出来，在同一图表中尽可能多的展现分析后的数据。

3. 项目解决思路

3.1. 总体思路

对于德鲁伊的翡翠梦境—APP 评论数据分析系统，我们分四个阶段寻求系统的解决思路：

- 1) 需求获取以及分析阶段，团队内部集中长时间讨论对项目要求的详细分析，明确项目的基本目标。并在团队讨论的收尾阶段邀请到某公司的项目经理来为我们进行 1 小时的项目指导。通过与外包项目指导老师（罗荣良老师）沟通交流，深入了解并确认需求。
- 2) 挖掘阶段，在需求分析的基础上，挖掘项目的核心价值，并对每一个价值点进行风险评估以及可行性分析，进而设计出可行的技术路线。
- 3) 整合阶段，充分整合分析出具体的技术路线，完成需求思维导图的制作。
- 4) 结题阶段，以需求思维导图为向导，完成最终方案的撰写。

3.2. 具体做法

德鲁伊的翡翠梦境团队使用了当下比较热门且适用于快速迭代开发的微框架 Spring-boot 并集成 mybatis、springMVC、thymeleaf 来完成后端开发。并在项目开发过程中使用了 Junit 测试。使用 python 语言对 AppStore 中的评论进行爬虫，以此获取大量数据并导入数据库。并对抓取的评论信息进行过滤、分词等处理，将分析后的数据以趋势线，词云图，饼图等相关图表手段形式直观地呈现给用户，具体功能如下图所示：

