目录

[1. 项目目标 2](#_Toc481027752)

[1.1. 项目背景 2](#_Toc481027753)

[1.1.1.公司背景 2](#_Toc481027754)

[1.1.2.技术背景 2](#_Toc481027755)

[1.1.3.业务背景 3](#_Toc481027756)

[1.2. 项目目标 3](#_Toc481027757)

[2. 项目服务模型 4](#_Toc481027758)

[2.1. 可行性服务模型 4](#_Toc481027759)

[2.1.1. 技术可行性 4](#_Toc481027760)

[2.1.2. 社会环境可行性 4](#_Toc481027761)

[2.1.3. 政策可行性 4](#_Toc481027762)

[2.2. 需求分析模型 4](#_Toc481027763)

[3. 项目价值 5](#_Toc481027764)

[3.1. 项目难点分析 5](#_Toc481027765)

[3.1.1. 评论关键信息的获取 5](#_Toc481027766)

[3.1.2. 提供评论反馈和项目跟进 5](#_Toc481027767)

[3.2. 项目优势分析 5](#_Toc481027768)

[3.2.1. 技术优势 5](#_Toc481027769)

[3.2.2. 应用优势 6](#_Toc481027770)

[4. 项目解决思路 6](#_Toc481027772)

[4.1. 总体思路 6](#_Toc481027773)

[4.2. 具体做法 6](#_Toc481027774)

# 项目目标

## 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。

2004年，虹软捕捉到手机相机市场蕴藏巨大潜力，开始专注于手机平台的影像处理和拍摄技术，成为世界上最早进入移动领域的传统影像软件公司。之后，随着主流手机平台的发布，虹软基于一向专注的多媒体、图像等领域开展了深度研发，技术成果随着巨量移动设备进入到消费者手中。并随后在竞争日益激烈的智能手机市场环境中，凭借在计算摄影与计算机视觉领域的研发领导地位，集中开发了一系列独特的技术和产品，稳固地帮助各大厂商、互联网公司建立了差异化的产品和良好的口碑，成为了他们的最佳拍档。

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。

### 1.1.2.技术背景

Viiision—APP评论数据分析系统是一个集数据爬取、分词技术、评论分析、机器学习、可视化技术等新一代信息技术为一体的多样性系统。致力于通过对所挖掘数据的处理和分析，将处理后的结果通过各种类型的图表方式直观地呈现给用户，使得用户能够方便快捷地获取所需信息。该系统根据用户分组为用户设计个性化的查看信息内容，用户也可根据自己所需对所查看内容进行更改或新增。

1、数据爬取

在我们日常生活中浏览网页的时候，往往会看到一些有价值的东西急于把它保存下载，却苦于下载需要注册亦或是要有积分要求，有的时候下载下来后也有损真实度。而用python可以来实现一个简单的爬虫功能，把我们想要的实时爬取到本地。本项目Viiision—APP评论数据分析系统主要是根据用户所需，实时地从各大应用商店中将APP总体信息和评论详情抓取过来，作为数据分析的重要来源。

2、评论分析

受互联网高速发展的影响，用户往往会在各大应用商店和论坛发表自己使用APP后的感受和想法，而这些感受来自用户的直接感受，对于企业来说相当重要。面对如此庞大的数据，企业都会手足无措，而Viiision—APP评论数据分析系统通过分词算法、关键词提取算法、垃圾评论过滤算法等技术对爬取的评论进行分析，将分析后的结果清晰地展现给用户和相关企业。

3、数据可视化

在这个大数据的时代，人人往往方案于单纯的数据呈现。近些年，数据可视化技术也在迅猛发展，数据可视化是利用计算机图形学的图像处理技术，将数据转换成图形或图像在屏幕上显示出来，并进行交互处理的理论、方法和技术。在本系统中，主要是将分析后的数据呈现给用户，如何充分地利用数据可视化技术就显得尤为重要。本项目组致力于通过不同样式的图表呈现使用户能够更加快捷、更加清晰地了解到相关信息。

### 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，显得十分的重要。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者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，也陆续出现了多款基于评论的分析系统，如：WeTest腾讯质量开放平台、ASO114、Cobub等，他们大多更加偏重于数据的丰富度，而对于数据的分析、呈现方式、结果反馈上不是特别看重。而本系统旨在给公司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。

## 项目目标

* **系统的适配性**

系统在电脑端和移动端都能正常访问，通过加强用户交互以带给用户更好的用户体验

* **用户使用的高效性**

实现用户在对相关数据进行导入时，占用资源较少，并同时保证查询页面依然可以正常访问，多人可以同时流畅地在线使用

* **用户查询的多样性**

**用户期望根据时间颗粒度、评论好坏程度、应用商店、评论关键字等有选择性地查看完整的评论详情**

* **数据呈现的直观性**

用户可以根据自己的需求方便快捷地从平台上获取相关的信息，且信息的呈现方式直观，易于被广大用户接纳

* **问题反馈的针对性**

企业内各部门职能不一，各部门期望获得有针对性的问题反馈，以便更好地提出后续的解决方案。

* **用户使用的个性化**

**用户期望可以对关注的APP进行增加和修改，希望对不同APP可以人为地增加分词并在下次查询时优先显示**

* **管理员管理的全面性**

**管理员期望能对用户进行增删改查并进行权限管理，能够方便地将本地数据或网上数据导入数据库中**

# 项目服务模型

## 可行性服务模型

### 2.1.1.技术可行性

数据分析型业务的痛点在于如何从海量的数据中找到用户所需的信息并以一种直观的方式展现给用户。因此对于数据源的要求是分类明确并且获取平台多样，对于呈现方式的要求是能让用户第一眼便看到自己想要的数据。由此可见对于信息的获取并筛选和筛选结果的呈现方式的选择尤为重要。

对于信息的获取和筛选，我们通过爬虫从各大应用平台获得相应的评论，并通过分词器和关键词提取算法分析出不同类别的词并提供给不同的部门，从而通过获得的信息改进自己的产品。

对于信息的展示，我们选用了Echarts库，通过动态可交互的柱状图，曲线图，地区分布图等来个性化地展现数据。

### 2.1.2.社会环境可行性

在当今的大数据时代，其战略意义不在于掌握庞大的数据信息，而在于对这些含有意义的数据进行专业化处理。换而言之，如果把大数据比作一种产业，那么这种产业实现盈利的关键，在于提高对数据的“加工能力”，通过“加工”实现数据的“增值”。

智能手机的发展使得所有的工具模块都已Application的形式存在，更甚包括蒂姆·库克在发布Apple TV时，也表示Apple TV的未来是APP。可见在数码设备上，APP的地位是越来越稳定的。在这样的条件下，对于APP及时的反馈就显得尤为重要。

Viiision的平台针对于用户对于APP全面的反馈进行筛选和总结，以图表的形式呈现，为APP的良性发展提供意见。

### 2.1.3.政策可行性

2015年7月4日，国务院印发《国务院关于积极推进“互联网+”行动的指导意见》。政府大力支持“互联网+”产业。Viiision通过数据可视化为移动应用的良性发展开辟新的道路，为“互联网+”奉献自己的一份力。

## 需求分析模型

企业的目的在于通过大量关于用户下载和评论的操作，经过数据挖掘来获取信息，从而为自己已有的产品提供改进思路并为新的产品提供构思。

第一阶段：我们对企业提供的命题介绍文档进行了分析，企业的需求进行总结并把它汇总，提供给前端进行页面设计。

第二阶段：项目组成员进行会议，汇总出命题中的模糊点，通过E-mail向企业提问进行更明确地产品定位。

第三阶段：为了给企业每个部门提供更具有针对性的信息，我们对分词内容进行个性化定制。

# 项目价值

## 项目难点分析

### 评论关键信息的获取

本系统的核心功能为从海量评论数据中为用户提取有用信息。因此关键词提取效果的好坏决定了该系统的成功与否。针对该难点，我们采用基于TF-IDF算法作为关键词选取的主要算法，确保提取的关键词具有较高的代表性和区分性。

为不同需求用户提供差异化的分析服务也是提取关键信息时的难点之一。不同的部门之间的需求既存在差异点又有共通之处，在提供关键词分类时需要权衡需求上的差异与共性。

### 提供评论反馈和项目跟进

在为不同部门用户提供个性化的评论分析后，用户期望通过得到的分析结果对其自身产品做出改进。如何更有效得帮助其挖掘信息和后续的跟进是本小组将要着手解决的难点。我们期望通过为用户提供反馈和项目不同版本阶段的标记来解决该问题。

## 项目优势分析

### 技术优势

#### 爬虫技术实现海量数据的抓取

对于来自不同平台的数据进行数据解析及结构化存储。同时采用分布式爬虫技术实现并行爬取操作，提高数据获取效率。强大的爬虫框架实现针对于不同平台的自适应配置，保证用户通过简单配置平台地址就可获取平台评论数据的便捷性。

#### 准确可靠的评论数据分析算法

平台针对评论文本分析的不同阶段使用了不同的算法。通过基于Trie树结构的词典扫描生成有向无环图和基于动态规划查找最大概率路径的分词算法实现评论文本的准确分词。通过基于TF-IDF算法实现评论关键词提取算法，同时针对不同的用户需求面改进了分类词典，保证关键词提取的针对性。最后采用了去重和垃圾过滤算法解决评论数据分析的噪声问题。

#### 前端交互技术保障页面的流畅性和友好性

采取结构统一、色彩均衡的界面设计提高用户的感官体验。使用响应式的页面布局保证用户能在不同尺寸的设备上获取合适的页面。使用AJAX异步加载数据的方式避免重刷页面导致的用户体验下降。

提供形式多样的条件筛选方式，满足不同用户希望获取不同粒度、不同方面信息的交互需求。

#### 调用Echarts提供直观的统计分析功能

由于平台提供大量关于APP的包括评论量、下载趋势、评论数目、平台分布、地域分布等的数据的统计分析功能。这些统计信息引用后台数据，调用Echarts提供的API，通过各种图表的方式直观地显示，为用户关于该APP近况和评论分析结果提供直观的统计分析认知。

### 3.2.2.应用优势



#### 3.2.2.1.有针对性的用户权限管理平台

根据用户职能将用户分为运营部、项目部和开发部。管理员对这些部门进行管理，分配可以查看的分词内容，运营部可以查看活动类的分词，开发部可以查看功能类的分词，产品部可以查看功能类和问题类的分词，且可根据所需新增其他部门的分词优先查看。

#### 3.2.2.2.个性化的用户APP管理平台

用户可以打开左侧栏对关注的APP进行分组管理，也可新增相关APP，对各个APP，可以实时地添加相关分词，使得每个用户对于各个APP都有一个个性化的词库，且在查询时，优先显示优先级高的分词。

#### 3.2.2.3.及时的评论统计分析反馈功能

通过对评论进行分词处理，将关键词进行分类管理，例如稳定性问题、交互问题、逻辑问题等，交由不同的部门进行分析解决，并对同一部门中出现的问题，通过权重大小评定优先级，并以邮件的形式通知各部门，方便各部门按照紧迫程度进行问题求解。对版本更新或功能更新的时间点前后进行对比分析，将统计结果同样以邮件的形式通知相关部门。

#### 3.2.2.4.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，我们采用的图表主要有如下几种形式：1.词云图的形式，该图集某一时间段内热词排行榜、某一热词词频随时间变化趋势、不同日期的热词分布这三种信息为一体，且用户可以通过点击相关热词查看包含该热词的好评与差评评论。2.柱状图与曲线图的形式，将下载量和评论量这两个信息结合在一张图中，展现其按照时间的变化情况。3.地图的形式，将不同地区的APP信息展现在地图上，从而可比较在不同的地区，某一APP的信息分布状况。

# 项目解决思路

## 总体思路

对于Viiision—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. 分析阶段，通过对项目要求的详细分析和外包服务业务的背景学习，明确项目的基本目标。通过社会调研以及与发包商的沟通交流，深入了解需求。
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成本分析和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，提出宏观的系统架构。
4. 结题阶段，根据整合得到的成本分析，服务模型以及相应的技术路线设计系统最终完成解决方案的撰写。

提交解决方案，通过分析审核后进行项目的实施

## 具体做法

Viiision—APP评论数据分析系统主要面向企业内的各个部门，同时需要管理员定期管理后台，因此根据使用人群的不同将该系统分为用户平台和管理员平台，用户平台主要是通过图表的形式呈现给用户APP评论分析详情，管理员端主要是方便管理员对用户信息和数据库内容进行管理，具体功能如下图所示：



具体功能模块详情如下表所示：

|  |  |
| --- | --- |
| 使用者需求分析 | |
| APP概况浏览功能 | 使用者可以分时间颗粒度查看关注的APP评论近况，评论近况包括APP下载量变化、评论数增长变化、喜好程度或态度变化等 |
| APP版本比较功能 | 使用者希望直观地了解到APP版本更新前后的变化 |
| APP地区比较功能 | 能够通过图表直观地了解到不同地区对于APP的使用和反馈情况 |
| 不同应用商店比较功能 | 能够根据所需，了解到APP在不同的应用商店中的反馈情况 |
| 评论查询功能 | 使用者可以按照时间、应用商店、关键词等搜索条件，查询相应的评论详情 |
| 评论分词统计功能 | 用户期望通过词云图展示大家关注的内容并统计改词在不同地区的区域分布情况 |
| 评论内容反馈功能 | 希望将评论分析的结果以图表的形式展现出来，实时了解与掌握用户关注的内容与趋势，了解用户反馈的问题 |
| 评论分词编辑功能 | 使用者可以根据所需编辑评论的拆分分词内容，在下次查看时优先显示 |
| 个性化选择功能 | 希望能够方便快捷地对关注的APP进行管理，并且可为不同的APP添加不同的分词，实现个性化 |
| 问题反馈功能 | 使用者期望通过查看反馈内容了解用户需求，更改解决方案 |
| 移动端适配功能 | 用户希望能在电脑端和移动端正常访问，已达到随时随地查看的目的 |

|  |  |
| --- | --- |
| 管理员需求分析 | |
| 更改用户信息功能 | 能够方便、快捷、高效地对用户进行增删改和分组 |
| 管理用户权限功能 | 对不同的用户设置不同的权限，并分配可以查看的APP信息 |
| 评论本地导入功能 | 期望将整理好的评论详情以Excel的格式方便地的导入到系统中 |
| 评论自动导入功能 | 期望通过配置获取地址、用户名和密码实现从网络平台自动获取数据 |
| 自动分词归类功能 | 能够自动对获取的评论内容进行解析，并实现分类 |
| 添加分词功能 | 期望人为地对不同的APP添加关注分词内容，并能优先展示 |
| 按颗粒展示功能 | 可以根据所需查看不同颗粒度下的评论详情 |