### APP评论数据分析系统

# 【命题企业介绍】

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。

注1：命题企业将对本项目获二等奖以上的优秀团队颁发奖金。

# 1、背景说明

## 【整体背景】

当今社会，随着智能机的普及，APP的开发和使用也是层出不穷。虹软作为一个走在新技术前沿的公司，希望能够开发出最贴合用户需求的功能。但是由于互联网无疆界，来自全球各地的人有着各自不同的需求，如何能够快速的了解并掌握世界各地用户的最新需求，是每个科技公司的迫切需求，虹软希望能够在最快的时间满足最新的需求。**我们希望能有一个系统，在获得海量用户的反馈基础上，能够较为快速的分析出不同地区，不同的人对于该APP的反馈。让公司决策者和开发者能够第一时间了解到最重要的信息。**

## 【公司背景】

虹软公司一直专注于影像和多媒体软件的技术公司。1994 年成立，为OEM制造商提供先进的应用软件、为电信业者与消费性电子产品公司提供基础建设解决方案。在研发或市场营销一直都是产业前瞻领导者。公司提供适用于数码相机、个人电脑、外设、移动终端设备的多媒体嵌入式软件产品以及消费电子固件方案。

## 【业务背景】

虹软公司是基于多媒体软件技术的一家公司，与世界上各大移动设备生产商有十分紧密的合作关系，为他们提供优秀的图形图像解决方案。不同的解决方案，可以衍生出不同的应用。而如何有效的提升应用的使用率并挖掘潜在需求，是面临的一个难题。

# 2、项目说明

## 【问题说明】

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应用的使用感受，这些感受因为来自直接客户的评价，因而十分珍贵。由于评论量日积月累非常庞大，且每个平台的用户在语言、用词等方面都会有较大的不同，如何在各种语言以及不用的用词习惯中，精确的获取到对企业有用的信息，则变得非常的重要。

企业需求如下：

* 能够把各个平台download下来的评论内容，方便的输入到该系统中。
* 系统能够自动进行语法分析，并进行分词后保存在数据库中。
* 能够查看具体的每一条评论信息，包含应用商店、日期等评论相关属性。
* 希望通过该系统，能够查看近期一段时间内（如近一周/月/年）的反馈情况。
* 能够区分并展示多款不同的APP表现情况。
* 能够区分APP在不同的应用商店反馈情况。
* 反馈内容，包括好评、差评、中评的态度内容，以及反馈的问题。
* 能够通过趋势线，词云图，饼图等相关图表手段表现出以上内容。

通过开发一套评论分析系统，有助于让企业及时的获取到更为精准的产品使用信息。在出现问题时可以及时纠正，也可以在市场商机出现时及时把握。通过人工拆分数据，或者算法自动拆分数据，再通过海量数据的整合分析，可以整理出在某一阶段，该产品客户的使用倾向与喜好程度。

## 【用户期望】

用户主要包括两类：使用者（决策者或开发者）、管理员。

使用者期望能够通过该系统，清晰的了解到用户在某一时期对该产品某个方面的反馈，如该产品的下载趋势、评论好坏趋势、以及建议统计等信息。

管理员希望能够将不同的平台的评论内容（一般为Excel或者从平台数据网站自动获取），较为方便的导入到该系统中。该系统能将相应的拆分分词保存在数据库中，并在打开该系统后，能够较为方便的向使用者展示所需的统计信息。

展示的内容期望如下：



结构图（图1）

**使用者期望：**

* 首页说明：
* 能够根据用户设置，展示最关注的APP评论近况。近况内容主要包括：近一周/月/年的下载量变化，近一周/月/年的评论数增长变化、喜好程度或态度变化。
* 统计分析：
  + 打开默认显示关注的APP近况（可能有两个或以上），可以查看特定时间内的APP近况（近一周/月/年的评论数增长变化，近一周/月/年的喜好程度或态度变化等内容）。
  + 近况包括但不限于，单位时间内评论数增长趋势、喜好度变化情况、打分趋势，同时还需显示相同APP不同版本间对应的近况。
  + 通过词云、趋势图等形式展示最近大家都关注什么。同时统计该词在不同的地区区域分布情况。
* 评论详情：
* 可以按时间、内容、地区、分数等搜索条件，查询相应的评论详情。并可以通过该页面，编辑该评论的拆分分词内容。
* 用户设置：
* 可以设置用户密码、关注的APP、以及APP类管理等功能。
* 增对不同的APP，可以添加“关注分词内容”。添加后的分词，应能优先分析展示。

每月下载量示意图（图2）

**管理员期望：**

* 用户管理
* 可以对用户进行增删改、分组
* 对不同的组，可以分配可以查看的APP信息
* 数据导入：
* 管理员可将整理好的Excel内的评论内容，方便的导入到系统中。Excel内容主要包括，评论内容，打分情况（0-5分），区域，语言等信息。
* 为了减少工作量，还期望系统能够自动从网络平台获取数据，管理员只需要配置获取地址，及用户名和密码。并可以自动对获取的内容进行分词解析，自动归类。（如评语：我非常喜欢这个软件，可以解析出“喜欢”这个态度，以及“非常”这个程度）
* 分析管理
* 增对不同的APP，可以添加“关注分词内容”。添加后的分词，应能优先分析展示。
* 可以设置默认展示的颗粒度（如是按年展示还是按月展示）

# 3、任务要求

* 项目要求
  1. 系统需要能够在电脑端和移动端正常访问，并针对移动端设备作适配优化。
  2. 导入数据时不能占用太多资源，需保证查询页面依然可以正常访问。
  3. 能够满足多人（20以上）同时在线使用和查看而没有明显的卡顿。
* 技术要求
  1. 开发平台：支持Windows平台或Linux平台
  2. 客户端：支持主流浏览器直接访问，如IE、Chrome、Firefox、Safari等
  3. 开发语言：后端语言任选，前端JS、HTML等
  4. 应用服务器：Tomcat6.0，Nginx, IIS
  5. 数据库服务器：SQL Server、Mysql、Oracle或Mongodb等
  6. 考虑系统的针对性，安全性，易用性，
  7. 系统具有良好的扩充性