**初步需求调查分析**

**一、背景介绍**

知乎是中文互联网中知识量级问答较为全面的社区。不同领域的相关专业人士在这里对其研究领域的问题进行分析解答，方便人们获取有关信息。尽管有着专业的回答，但也充斥者不少低质量和无关回答。同时存在者用户阅读过某位作者的回答后想了解这个作者的相关其他回答，或者了解同类型作者难的问题。因此我们希望可以制作相关工具通过对知乎数据的排序与筛选提供给用户更精确数据和相关人士系列数据。

1. **欲解决问题**

对知乎用户进行统计并分析, 同时可以获取对应用户的详细信息、问题、回答、专栏等数据。对用户进行相关领域的排序和系统整理。

1. **软件功能描述**
2. 对知乎用户进行其领域内的排序，将高质量的知乎用户及其回答提供给用户
3. 对单个知乎用户的所有回答进行爬取，提供给用户。
4. **系统的组成与部署**

程序初步设定为爬虫、后端处理、前端三个部分。爬虫可采用python script或者java webmagic进行知乎数据爬取，两种爬取方式，目前仅速度之间存在差距。

存储初步计划采用lucence进行数据存储，目前设定为用户id、用户回答、用户关注等数据。采用lucence，自身做了优化数据存储等操作，数据量大可以更加快速的索引。

前端目前有两个方案，第一个为vue方案，采用vue进行前端设计，提供给用户交互页面。第二方案为谷歌插件，将最终交互及结果通过插件进行展示。第一方案页面会更加美观方便，但是存在者单独页面，不方便用户使用的情况。第二方案界面美观程度不如第一个界面，但是在使用知乎的时候对用户来将更加的方便。