**文档编号：*百万级知乎用户数据分析与搜索系统* – SRS – *2.0***

***百万级知乎用户数据分析与搜索系统***

**软件需求规格说明书**

**日期：2021年10月17日**

**文档变更历史记录**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 变更日期 | 变更人员 | 变更内容详情描述 | 变更后的版本号 |
| 1 | 2021\9\28 | 全员 | 撰写了软件需求规格说明书初稿 | V1.0 |
| 2 | 2021\10\10 | 田力涛 | 更新软件UI图 | V1.1 |
| 3 | 2021\10\11 | 全员 | 更新软件功能 | V1.2 |
| 4 | 2021\10\17 | 全员 | 更新软件需求模型 | V2.0 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**目录**

[1. 引言](#_Toc8397)

[1.1 编写目的](#_Toc2032)

[1.2 读者对象](#_Toc3477)

[1.3 软件项目概述](#_Toc8237)

[1.4 文档概述](#_Toc16319)

[1.5 定义](#_Toc21806)

[1.6 参考资料](#_Toc2528)

[2. 软件的一般性描述](#_Toc22652)

[2.1软件产品与其环境之间的关系](#_Toc897)

[2.2假设与前提条件](#_Toc26392)

[3. 软件功能需求描述](#_Toc21198)

[3.1 软件功能概述](#_Toc16102)

[3.2 软件需求的用例模型](#_Toc21730)

[3.3 软件需求的分析模型](#_Toc21730)

[3.3.1“用户进入首页”的用例描述](#_Toc3155)

3.3.2“用户搜索”的用例描述

3.3.3“用户进入图表页面”的用例描述

3.3.4“用户选择图表格式”的用例描述

3.3.5“用户筛选图表信息”的用例描述

3.3.6“用户查看排行榜”的用例描述

[4. 其它软件需求描述](#_Toc10848)

[4.1 性能要求](#_Toc9359)

[4.2 设计约束](#_Toc12177)

[4.3 界面要求](#_Toc26380)

[4.4 进度要求](#_Toc10119)

[4.5 交付要求](#_Toc8195)

[4.6 验收要求](#_Toc31930)

[5. 软件原型](#_Toc30787)

# **1. 引言**

## 1.1 编写目的

1)本文档的目的在于方便用户、分析人员和软件设计人员进行理解和交流。用户通过需求规格说明书在分析阶段即可初步判定目标软件能否满足其原来的期望，但是本文档主要是作为设计人员的软件开发的基本出发点和系统维护人员发现和添加新功能需求的基础，也是维护人员的技术支持文档之一。

2)本文档支持目标系统的确认。软件开发目标是否完成不应由系统测试阶段的人为因素决定，而应根据需求规格说明书中确立的可测试标准决定。

3)本文档控制系统进化过程。在需求分析完成后，如果用户追加需求，那么需求规格说明书将用于确定追加需求是否为新需求。如果是，开发人员必须针对新需求进行需求分析，扩充需求规格说明书，进行软件再设计。

## 1.2 读者对象

用户，分析人员，软件设计人员，项目管理人员。

## 1.3 软件项目概述

* 项目名称: 百万级知乎用户数据分析与搜索系统
* 用户单位: 知乎网友
* 开发单位:北京信息科技大学计算机学院19级软件工程难上加男小组
* 软件项目的背景和大致功能：

大数据下的时代，信息繁杂冗余，各类网站上的信息复杂多样，为了满足日常生活需求，信息被分类方便人们快速查询。知乎是当先访问量比较高的网站，但是缺少对一些信息的分类，这样使得在使用时会有一些不便利。例如查找某个用户的一些回答，或者相似回答时就会比较难以查询。因此在这样的情况下，我们希望解决这种信息查询方面的缺陷。

我们根据需求，设计了“百万级知乎用户数据分析与搜索”这一项目，用于

补足知乎在检索方面的不足。其利用爬虫抓取知乎用户的数据并进行分析，处理，将其对用户可视化

## 1.4 文档概述

1）软件的一般性描述部分。它包括软件产品与其环境的关系、软件受到的限制和约束以及软件开发前的假设与前提条件。

2）功能需求描述部分。它主要分为系统的划分，软件各子系统的功能，设计约束和性能、界面、交付、验收四个方面的要求。

3）其它软件需求描述部分。它包括性能要求、设计约束、界面要求、进度要求、交付要求和验收要求。

4）软件原型。主要设计了软件运行不同功能时的界面，比如用户登录界面。

## 1.5 定义

无

## 1.6 参考资料

[1].软件工程.齐治昌，谭庆平，宁洪.北京:高等教育出版社，2012

[2].需求分析与设计.马素霞译.北京:机械工业出版社，2009

[3].面向使用的软件设计.刘正捷译.北京:机械工业出版社，2011

[4].第一次民间版知乎用户分析报告. 2016

[5].民间版知乎分析报告第二期——赞同背后的秘密. 2018

# 

# **2. 软件的一般性描述**

### 2.1软件产品与其环境之间的关系

本软件使用Python语言从相应知乎中获取用户数据并对其进行数据挖掘, 处理后的数据利用MySQL数据库进行保存, 并且利用vue+echarts搭建的前端以及springboot+mybits+maven搭建的后端对其进行可视化处理, 并展示在网页上。

表 2-1 : 软件系统组成部分及其运行环境

|  |  |
| --- | --- |
| **软件系统组成部分** | **运行环境** |
| 爬虫 | Python |
| 数据库 | MySQL |
| 前端 | Vue + Element + ECharts |
| 后端 | Sprint Boot + MyBatis + Maven |

### 2.2限制与约束

### (1) 功能完整性限制:

### 软件应保证其主要功能完整, 且其功能可以稳定运行。

### (2) 功能响应时间限制:

软件系统应保证有很强的健壮性, 不能因为用户并发使用而导致系统崩溃, 且系统建成后应满足以下性能要求：

平均响应时间: 50用户小于3秒; 200用户小于30秒

CPU平均使用率小于80%

(3) 可靠性限制:

软件系统应能够保持长期稳定运行。

(4) 环境约束:

软件开发环境应与上述环境一致。

(5) 时间约束:

本软件应于2021年11月30日前完成该软件的主要开发任务及工作, 保证该软件达到上述限制标准。

### 2.3假设与前提条件

开发该软件时, 我们默认用户熟悉基本的网页操作, 管理员能够熟练地进行后台调试及维护工作。

# **3. 软件功能需求描述**

### 3.1 软件功能概述

1. 对百万级数据进行爬取、清理, 并根据用户群的性别、 年龄层次、 喜好分区等种类对清洗后的数据进行分析、统计及可视化，做到短周期内同步更新，确保数据的真实性以及有效性。（级别1）
2. 对分析数据进行图表可视化展示（级别2）
3. 根据知乎用户的关注数量, 赞同数量对用户进行排名, 供使用者查看（级别2）
4. 对知乎中用户的详细数据（应包含其回答，视频，专栏等数据）提供下载功能，供使用者进行下载（数据会以文件包的形式提供给用户）（级别2）
5. 简单知识图谱构建（级别3、待定功能）

**注: 级别数字越小, 表示其优先级越高**

### 3.2 软件需求的用例模型

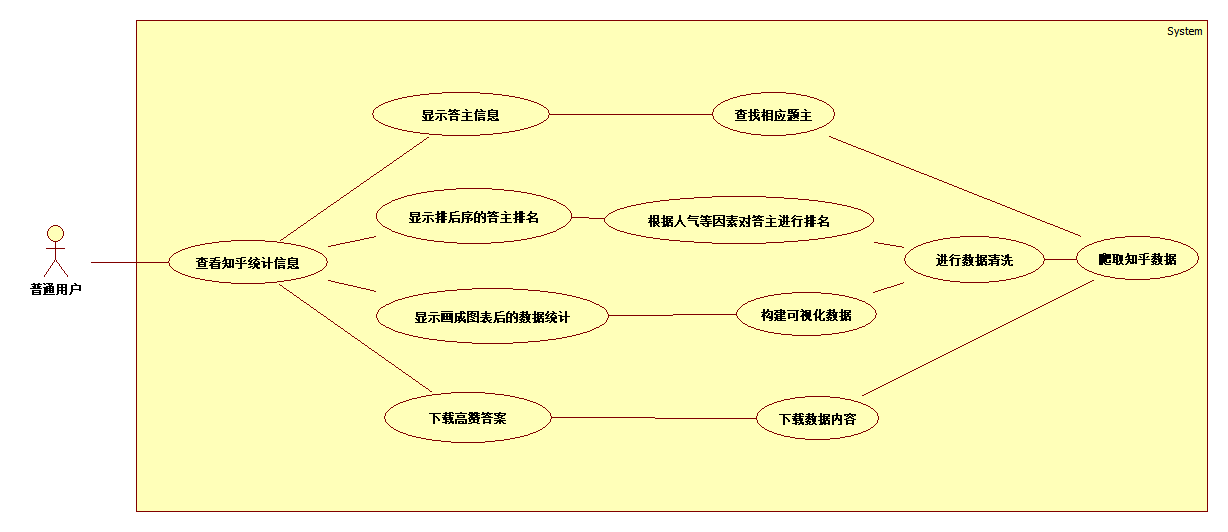


图3.1 用例图

### 3.3 软件需求的分析模型

***1、查看知乎统计信息***

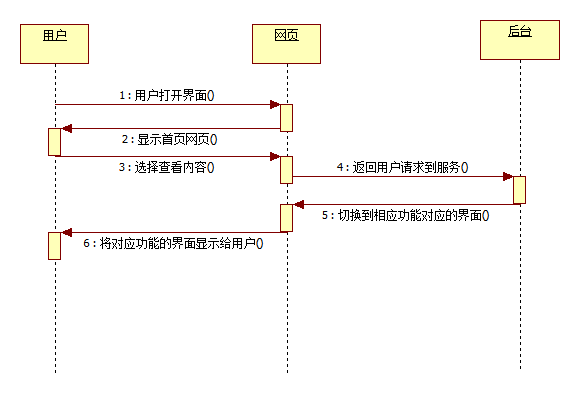


图3.2 系统顺序图-查看知乎统计信息

表3-1 用例描述-查看知乎统计信息

|  |  |
| --- | --- |
| 用例名 | 查看知乎统计信息 |
| 用例描述 | 用户进入网页查看统计信息 |
| 参与者 | 用户 知乎 |
| 过程 | 1.用户打开界面  2.显示首页网页  3.选择查看内容  4.返回用户请求到服务  5.切换到相应功能对应的界面  6.将对应功能的界面显示给用户 |

1. ***显示答主信息***

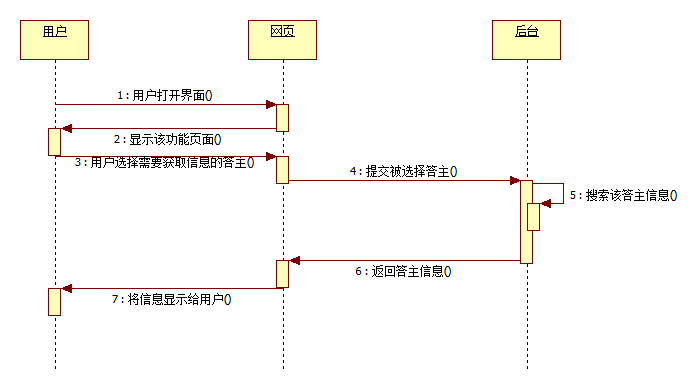


图3.3 系统顺序图-显示答主信息

表3-2 用例描述-显示答主信息

|  |  |
| --- | --- |
| 用例名 | 显示答主信息 |
| 用例描述 | 显示答主信息 |
| 参与者 | 用户 知乎 |
| 过程 | 1.用户打开界面  2.显示该功能页面  3.用户选择需要获取信息的答主  4.提交被选择答主  5.搜索该答主信息  6.返回答主信息  7.将信息显示给用户 |
|  |  |

1. ***显示排序后的答主排名***

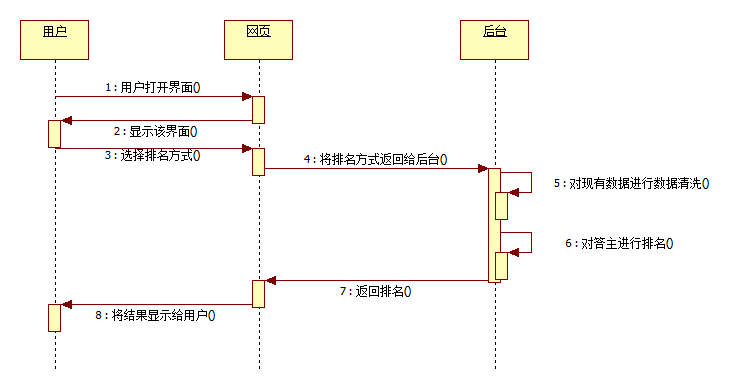


图3.4 系统顺序图-显示排序后的答主排名

表3-3 用例描述-显示排序后的答主排名

|  |  |
| --- | --- |
| 用例名 | 显示排序后的答主排名 |
| 用例描述 | 显示排序后的答主排名 |
| 参与者 | 用户 知乎 |
| 过程 | 1.用户打开界面  2.显示该界面  3.选择排名方式  4.将排名方式返回给后台  5.对现有数据进行数据清洗  6.对答主进行排名  7.返回排名  8.将结果显示给用户 |

1. ***下载高赞答案***

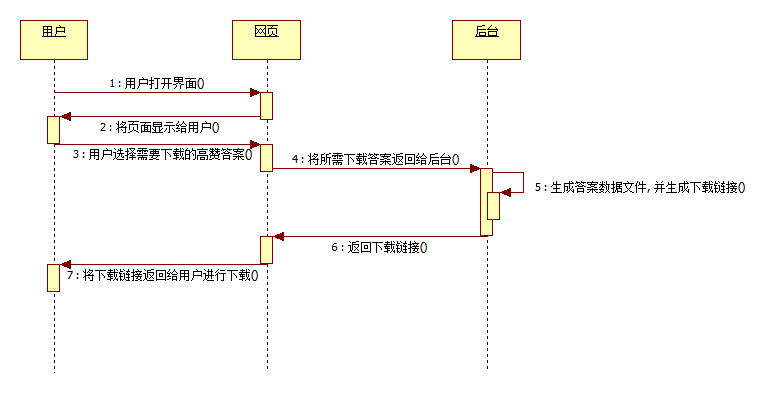


图3.5 系统顺序图-下载高赞答案

表3-4 用例描述-下载高赞答案

|  |  |
| --- | --- |
| 用例名 | 下载高赞答案 |
| 用例描述 | 下载高赞答案 |
| 参与者 | 用户 知乎 |
| 过程 | 1.用户打开界面  2.将页面显示给用户  3.用户选择需要下载的高赞答案  4.将所需下载答案返回给后台  5.生成答案数据文件, 并生成下载链接  6.返回下载链接  7.将下载链接返回给用户进行下载 |

1. ***显示化成图表后的数据统计***

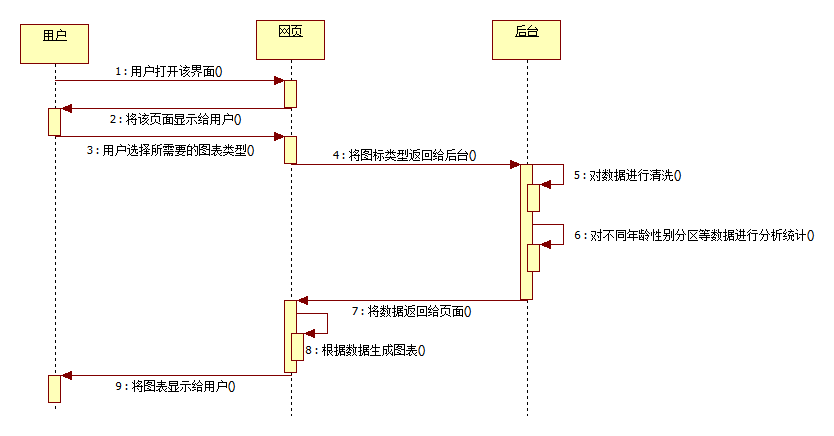


图3.6 系统顺序图-显示化成图表后的数据统计

表3-4 用例描述-显示化成图表后的数据统计

|  |  |
| --- | --- |
| 用例名 | 显示化成图表后的数据统计 |
| 用例描述 | 显示化成图表后的数据统计 |
| 参与者 | 用户 知乎 |
| 过程 | 1.用户打开该界面  2.将该页面显示给用户  3.用户选择所需要的图表类型  4.将图标类型返回给后台  5.对数据进行清洗  6.对不同年龄性别分区等数据进行分析统计  7.将数据返回给页面  8.根据数据生成图表  9.将图表显示给用户 |

# **4. 其它软件需求描述**

### 4.1 性能要求

### (1) 功能完整性要求:

### 软件应保证其主要功能完整, 且其功能可以稳定运行。

### (2) 功能响应时间要求:

软件系统应保证有很强的健壮性, 不能因为用户并发使用而导致系统崩溃, 且系统建成后应满足以下性能要求：

平均响应时间: 50用户小于3秒; 200用户小于30秒

CPU平均使用率小于80%

(3) 可靠性要求:

软件系统应能够保持长期稳定运行。

### 4.2 设计约束

(4) 环境约束:

软件开发环境应与上述环境一致。

(5) 时间约束:

本软件应于2021年12月31日前完成该软件的主要开发任务及工 作, 保证该软件达到上述限制标准。

### 4.3 界面要求

### 设计相应的网页界面, 要求数据的展示应进行可视化分析, 有联系的功能尽量展示在同一页面中, 针对数据可视化部分进行美化。

### 4.4 进度要求

开发者应于2021年10月31日前给出该程序的软件原形, 并在同年12 月31日完成软件的全部功能和开发工作, 完成验收与交付。

### 4.5 交付要求

### (1) 程序的前端、后端、爬虫以及数据库的全部软件代码。

(2) 程序设计规格说明书的电子和纸质文档。

(3) 使用说明书的电子和纸质文档。

### 4.6 验收要求

(1) 在验收过程中, 要求程序的整个系统正常运行, 运行过程中无任何bug, 在用户进行违规操作时弹出提醒且不会导致系统崩溃。

(2) 要求整个软件各个部分优化完毕, 不存在编码过程中遗留的代码问题导致用户体验受到影响。

# **5. 软件原型**

1）搜索UI图



图5.1 搜索UI图

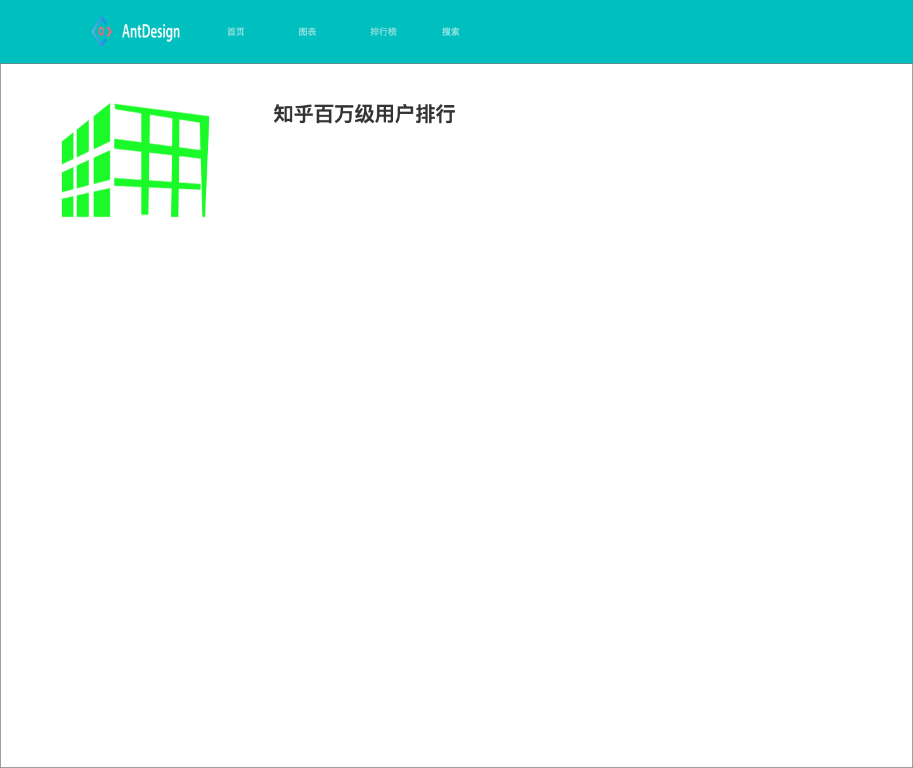
2）排行UI图

图5.2 排行UI图

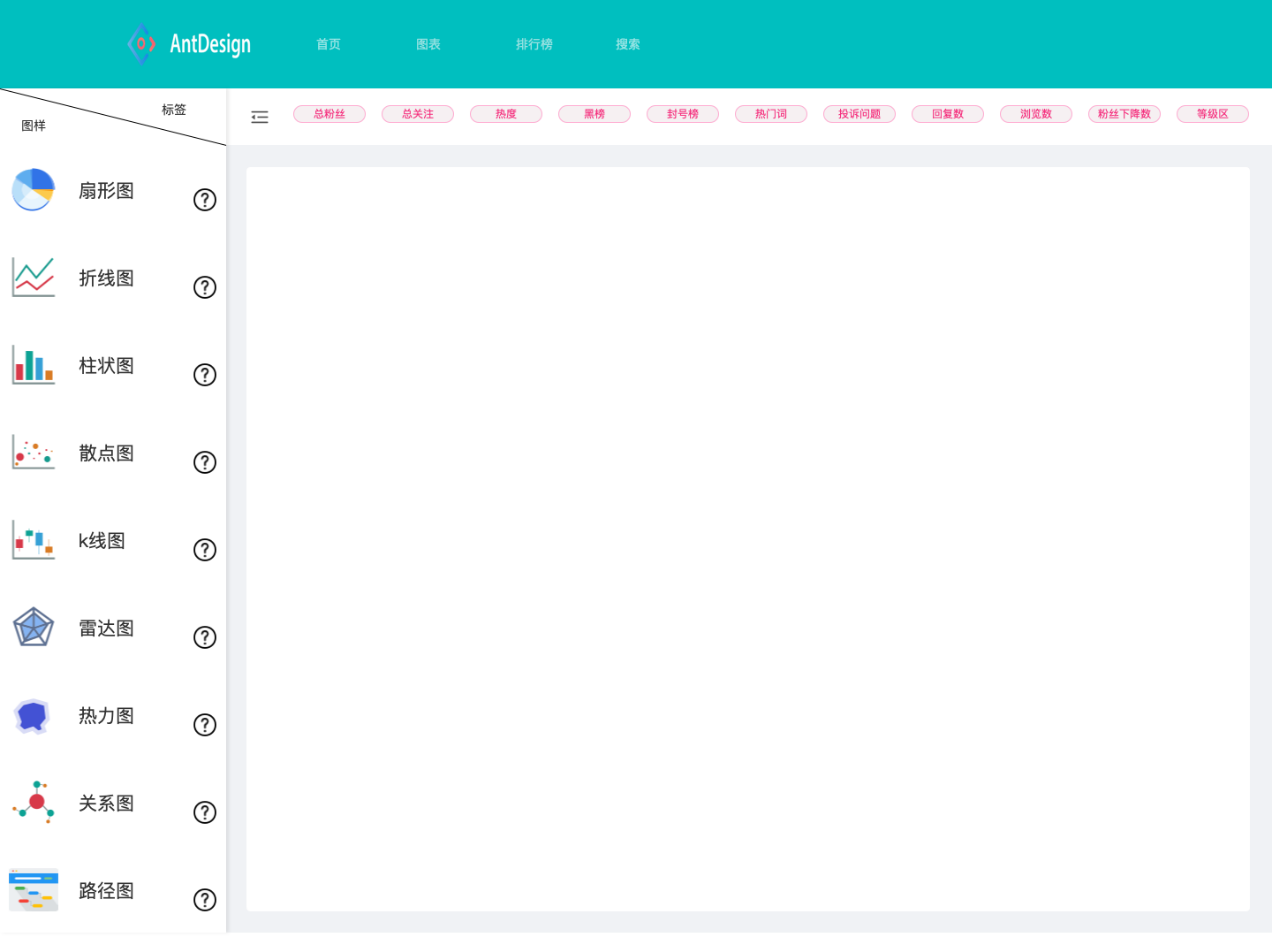
3）图标UI图

图5.3 图标UI图

4）用户详细内容UI图

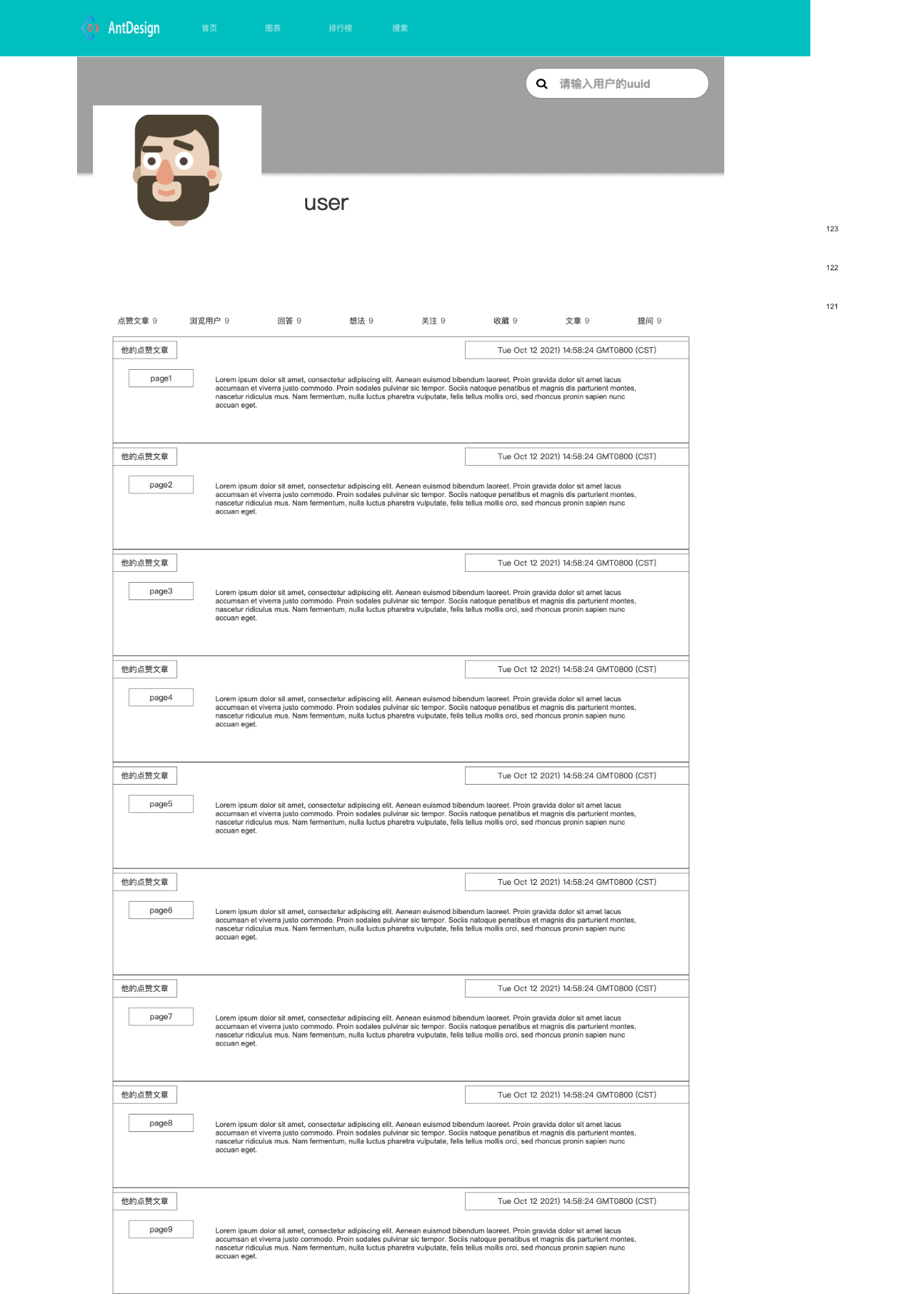


图5.4 用户详细内容UI图

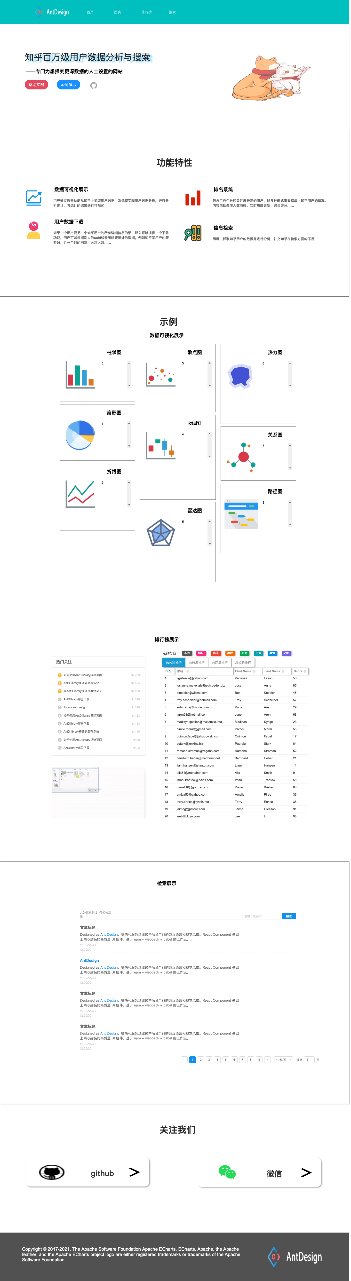
5）首页UI图

图5.5 首页UI图