**系统需求说明书**

**项目名称：头条助手**

**小组成员：谢黎颖，李智坚，蓝康，彭鼎，余捷**

**版本号：2.0.1**

**版权所有：第五组有限公司**

**完成日期：4月9日**

目录

[1、引言 1](#_Toc511051521)

[1．1编写目的 1](#_Toc511051522)

[1．2背景 1](#_Toc511051523)

[1．3定义 1](#_Toc511051524)

[1．4参考资料 1](#_Toc511051525)

[2、任务概述 2](#_Toc511051526)

[2．1目标 2](#_Toc511051527)

[2．2用户的特点 2](#_Toc511051528)

[2．3假定和约束 2](#_Toc511051529)

[3、需求规定 2](#_Toc511051530)

[3. 0 业务流程分析 2](#_Toc511051531)

[3．1对功能的规定 4](#_Toc511051532)

[R1.1网络爬虫 4](#_Toc511051533)

[R1.2主页个人新闻推荐 5](#_Toc511051534)

[R1.3综合新闻排序 5](#_Toc511051535)

[R1.2图表可视化 5](#_Toc511051536)

[3．2对性能的规定 6](#_Toc511051537)

[3．2．1精度 6](#_Toc511051538)

[3．2．2时间特性要求 6](#_Toc511051539)

[3．2．3灵活性 6](#_Toc511051540)

[3．3 数据分析 7](#_Toc511051541)

[3．5数据管理能力要求 9](#_Toc511051542)

[3．6故障处理要求 9](#_Toc511051543)

[3．7其他专门要求 9](#_Toc511051544)

[4、运行环境规定 10](#_Toc511051545)

[4．1设备 10](#_Toc511051546)

[4．2支持软件 11](#_Toc511051547)

[4．3接口 11](#_Toc511051548)

[4．4控制 11](#_Toc511051549)

# 1、引言

## 1．1编写目的

这里编写的“头条助手系统”的需求规格说明书，主要是为了明确项目的需求，并以其作为项目计划、概要设计、系统测试等的一个依据，同样也为参与此项目开发的成员、提供一个参考文档。本文档需要软件开发确认系统的需求是否准确、全面。

此规格说明书的读者对象范围包括：本项目的所有开发人员，测试人员。

## 1．2背景

头条助手系统是由第五小组同学提出的为了解决如今的新闻网站不能很好的利用不能迎合用户的新闻阅读习惯，较差的用户体验，用户不能体验到信息时代科技所带来的新鲜体验和简单快捷，从而失去阅读新闻的兴趣和习惯等问题提出的一个解决方案。所应用到的网络为Internet，该系统主要是从其他新闻网站上爬取相关数据进行处理再发布。

## 1．3定义

列出本文件中用到的专门术语的定义和外文首字母组词的原词组。

## 1．4参考资料

列出用得着的参考资料，如：

a本项目的经核准的计划任务书或合同、上级机关的批文；

b属于本项目的其他已发表的文件；

c本文件中各处引用的文件、资料、包括所要用到的软件开发标准。列出这些文件资料的标题、文件编号、发表日期和出版单位，说明能够得到这些文件资料的来源。

# 2、任务概述

## 2．1目标

信息社会时代，每个人都能随时随地地浏览新闻，已经逐渐成为人们日常生活中一部分，时事新闻已经越来越得到人们的关注。新媒体平台可以充分满足用户的新闻阅读需求，用最适合最便捷的方式，来营造最佳的用户体验，节约用户的阅读时间，提供精心细致而有全面的阅读服务。本系统可以显示通过爬虫后经过综合排序的和可视化的图表，还有可以在主页推荐新闻。

## 2．2用户的特点

主要针对的是14岁以上的新闻阅读爱好者，这类人群关心周围发生的新闻事件，渴望与外界世界取得密切联系。追求更全面的知识存储和高品质的生活。面向的人群不需要太多要求，只要能掌握基本的移动设备和使用和互联网知识即可。

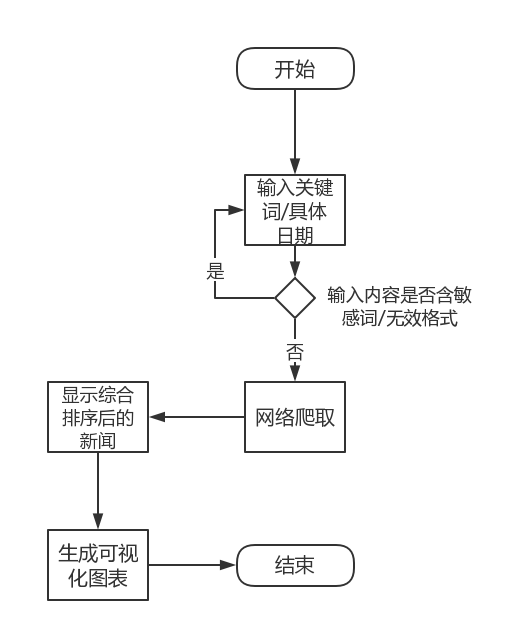
## 2．3假定和约束

本系统开发的周期为16周。

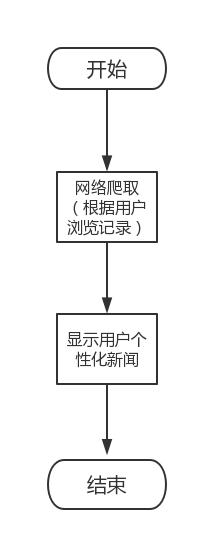
# 3、需求规定

## 3. 0 业务流程分析

1.输入内容综合排序新闻



2.推荐个性化新闻



## 3．1对功能的规定

#### 3.1.1对用户管理功能模块的概述

|  |  |  |
| --- | --- | --- |
| **标号（Ref#）** | **功能** | **备注** |
| R1.1 | 网络爬虫 | 从网上爬取相关新闻 |
| R1.2 | 主页个人新闻推荐 | 可以为用户推荐新闻 |
| R1.3 | 综合排序新闻 | 可以根据爬取的数据进行排序 |
| R1.4 | 图表可视化 | 可以用图表显示信息 |

|  |  |  |  |
| --- | --- | --- | --- |
| 功能 | 输入 | 处理 | 输出 |
| 用户输入关键词，日期 | 输入一些 | 在数据库中查找输入的省份是否存在，然后根据输入的省份查询数据库中该省份下的所有大学，返回查询结果 | 输出大学名单 |
| 网络爬虫 | 输入大学名称或直接在大学名单中选择大学 | 在数据库中查找输入的学校名称是否存在，然后将该大学的招生简章调出，返回查询结果 | 输出大学招生简章 |
| 显示排序后的新闻 | 输入大学名称或直接在大学名单中选择大学 | 在数据库中查找输入的学校名称是否存在，然后将该大学的入学指南调出，返回查询结果 | 输出大学入学指南 |
| 图标可视化 | 输入大学名称或直接在大学名单中选择大学 | 在数据库中查找输入的学校名称是否存在，然后将该大学的历年校线调出，返回查询结果 | 输出大学历年校线 |
| 查看专业列表 | 输入大学名称或直接在大学名单中选择大学 | 在数据库中查找输入的学校名称是否存在，然后将该大学的专业列表调出，返回查询结果 | 输出专业列表 |
| 查看高考咨询 | 输入省份（默认全国） | 在数据库中查找输入的省份是否存在，然后根据输入的省份查询数据库中该省份下的所有高考资讯，返回查询结果 | 输出高考资讯 |
| 高考查分 | 输入考生信息（准考证号和密码） | 验证考生信息，通过获取考生信息获取权限，然后访问官方的高考查分系统，将数据用网络爬虫爬取下来，返回查询结果 | 输出分数 |

#### R1.1网络爬虫

|  |  |
| --- | --- |
| **1、描述** | 根据用户输入爬取数据 |
| **2、参与者** | 系统管理员，用户 |
| **3、前置条件** | 用户输入关键词 |
| **4、结果** | 爬取数据成功或失败 |
| **5、基本流程** | 1、用户输入关键词  2、系统根据所写程序执行相关操作  3、将数据存入数据库中 |
| **6、异常** | 如果爬取不到数据，显示爬取错误 |
| **7、数据格式** | 字符串 |
| **8、返回数据** | 字符串 |
| **9、补充** | 关键词的字段必须合法，否则提示“请输入有效字段” |

#### R1.2主页个人新闻推荐

|  |  |
| --- | --- |
| **1、描述** | 系统根据用户的相关浏览历史进行新闻推荐 |
| **2、参与者** | 系统管理员 |
| **3、前置条件** | 用户进行新闻浏览 |
| **4、结果** | 成功显示推荐页面 |
| **5、基本流程** | 1、用户访问新闻页面  2、系统根据用户的浏览进行分析  3、系统推送推荐页面非用户 |
| **6、异常** |  |
| **7、数据格式** | 字符串 |
| **8、返回数据** | 无 |
| **9、补充** |  |

#### R1.3综合新闻排序

|  |  |
| --- | --- |
| **1、描述** | 系统对爬取的数据进行排序 |
| **2、参与者** | 系统 |
| **3、前置条件** | 成功爬取数据 |
| **4、结果** | 排序成功或失败 |
| **5、基本流程** | 1、系统成功爬取数据  2、系统对数据通过日期、热度进行排序  3、显示排序后新闻 |
| **6、异常** | 无 |
| **7、数据格式** | 字符串 |
| **8、返回数据** | 无 |
| **9、补充** | 无 |

#### R1.2图表可视化

|  |  |
| --- | --- |
| **1、描述** | 系统对爬取的数据进行图表化显示 |
| **2、参与者** | 系统 |
| **3、前置条件** | 成功爬取数据 |
| **4、结果** | 用图表来显示结果 |
| **5、基本流程** | 1、系统成功爬取数据  2、系统对数据通过日期、热度进行图表可视化  3、显示图表 |
| **6、异常** | 无 |
| **7、数据格式** | 字符串 |
| **8、返回数据** | 无 |
| **9、补充** | 无 |

#### 3.1.2关于数据的解释

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名 | 类型 | 说明 | 组成 |
| 用户输入关键词 | Nvarchar(20) | 实时热点和七日关注类为短期关键词  热门搜索和世说新词为长期关键词。 |  |
| 浏览历史 | text | 一个用于显示用户近七日浏览标题的文本文件 | 各个新闻热点的标题 |
| 用户输入的日期 | Datatime | 精准到年月日的日期输入 | 年月日 |

**短期热搜关键词：**

本类事件、人物或者话题出现快，但是也消失快。例如：明星事件，一部新出的电影，一个新出的电视剧，比如现在的港台的富贵门，韩国的创造情缘，国内的蜗居。

**长期热搜关键词：**

本类事件、人物或者话题比较长久，余音绕梁，经久不衰，在相当长的一段时间内都属于比较热的搜索关键词，不会轻易改变。比如：中国移动，淘宝等等。

## 3．2对性能的规定

### 3．2．1精度

1、在执行关键词搜索的时候，不允许出现因为程序的原因导致网络爬虫的失败，也不允许发生重复爬取的情况；

2、在执行排序操作的时候，不允许因为程序的原因发生数据丢失、排序错误的情况；

3、图表的显示也要求保持对应的准确性；

### 3．2．2时间特性要求

在单用户执行新闻浏览的时候，在运行环境规定的条件下，单次操作的响应时间要求在2秒钟之内。

返回100行以内的新闻，单次操作的响应时间要求在2秒之内。

多人操作时候，时间和相应的要求同上。

### 3．2．3灵活性

A、操作方式

程序在通常的应用环境下使用鼠标和键盘进行输入和输出操作，对于执行按钮，通常使用鼠标的点击完成，但是，界面要求全部支持键盘的定位操作（在不安装鼠标的计算机上，也能够使用该系统）。

B、运行环境

程序在通常的条件下，在Win10/xp/7上安装运行系统要求能够在Win95的操作系统上安装和运行。

C、同其他软件的接口的变化

（不适用）

* 1. 精度和有效时限的变化

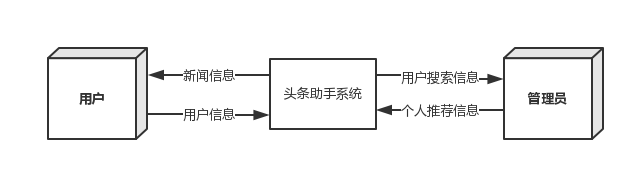
（不适用）

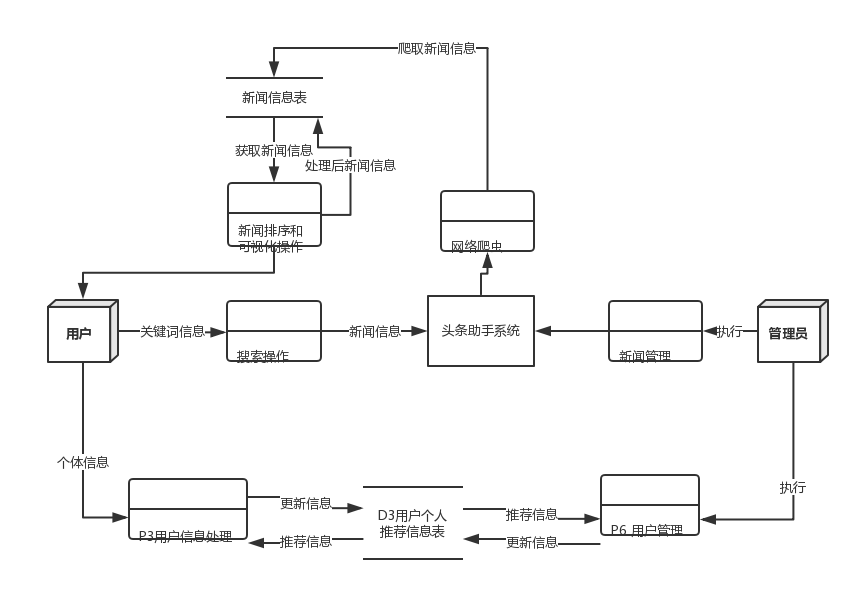
* 1. 计划的变化或改进

由于本系统的规模比较小，计划和进度的改变不影响到需要实现的需求。

## 3．3 数据分析

a.数据流图





数据字典：

1. 名称：新闻信息表

描述：记录网络爬取新闻的基本信息

来源：根据用户输入的关键词从网络上爬取

去处：头条助手中保存的新闻信息

组成：标题+导语+主语+ 背景+结语

1. 名称：用户个人推荐信息表

描述：记录用户新闻浏览后生成的个人推荐列表

来源：根据用户浏览新闻历史

去处：用户管理中的推荐信息表

组成：新闻种类+关键词

3．4输人输出要求

暂无

## 3．5数据管理能力要求

支持3万用户。支持GB级数据。

## 3．6故障处理要求

A、在操作成员输入一些不合理的数据的时候，能够进行一些合理的提示信息，不能因为输入错误而导致系统的错误，或者程序停止运行；

B、程序运行时，对服务器和网络通信故障能够识别并提示，当故障排除后，程序恢复正常运行；

C、数据库要求有灾难备份机制，以防止数据的全部丢失。

## 3．7其他专门要求

A、安全性

网络安全：能经受来自互联网的一般性恶意攻击。如病毒（包括木马）攻击、口令猜测攻击、黑客入侵等。因此，必须配备较强的网络安全防范、响应能力，为应用系统提供安全可靠的网络统计平台。

数据库安全：数据库级备份和恢复。数据库级用户进行角色和权限授权。使得在异常情况发生时，系统可以得以快速恢复，避免数据的丢失或将其影响降到最低限度。同样，要保证存储过程中数据不被非法访问和篡改。

数据保密：网络传递数据经过加密。需要保证数据在采集、传输和处理过程中不被偷窥、窃取、篡改。

应用系统的安全：通过对用户的身份鉴别，并实施相应的访问控制策略后，使用户只能完成得到系统授权的数据访问功能操作。用户只有经授权后才可以更新程序，避免因错误程序更新而影响系统的正常运行。

其他安全方面的要求，请参考合同的说明。

B、可扩充性

系统在开发完毕以后，应允许在后续的合同中进行功能的扩展或者功能的重新解释和实现。

C、健壮性

系统应该保证在一次开机三个月之内稳定运行，数据库在一些灾难事故中能够在系统安装好之后，两小时内恢复。

可靠性指标：在连续运行情况下，系统可靠性99.9999%。提供应用服务器集群技术和组件技术支持高可靠性和伸缩性。

D、易用性

易学易用、符合标准浏览器操作风格，丰富的联机帮助，人性化的操作界面。

E、可维护性

系统从设计上尽量考虑少做二次开发或者不做二次开发，直接通过系统配置搭建系统，从功能上具有通用性，易修改和扩展。软件开发使用组件技术，保证了可维护性高。

# 4、运行环境规定

## 4．1设备

**客户程序硬件要求：**

具有 Pentium III 处理器且满足以下要求的计算机：

最低 64 MB 内存

最小 2.1 GB 硬盘

鼠标

键盘

**服务器硬件需求：**

具有 Pentium III 处理器且满足以下要求的计算机：

最低 256MB 内存

最小 8 GB 硬盘

鼠标

键盘

## 4．2支持软件

**浏览器客户端：**

操作系统：Windows 98/NT /2000或更高版本

浏览器：IE 5.0以上或Netscape同等版本以上；

**数据库服务器软件：**

操作系统：Windows NT /2000 Server或更高版本

Web服务器： Tomcat5.0等；

应用服务器：

数据库产品：SQL Server 2000

## 4．3接口

本系统通过操作系统的支持对硬件进行操作。

系统采用HTTP和HTTPS协议支持远程用户对系统的访问。

## 4．4控制

说明控制该软件的运行的方法和控制信号，并说明这些控制信号的来源。