

1 引言	2
1.1 编写目的.....	2
1.2 背景.....	2
1.3 定义.....	2
1.4 参考资料.....	2
2 用途	3
2.1 功能.....	3
2.2 性能.....	4
2.2.1 精度.....	4
2.2.2 时间特性.....	4
2.2.3 灵活性.....	4
2.3 安全保密.....	错误!未定义书签。
3 运行环境	错误!未定义书签。
3.1 硬设备.....	错误!未定义书签。
3.2 支持软件.....	错误!未定义书签。
3.3 数据结构.....	错误!未定义书签。
4 使用过程	错误!未定义书签。
4.1 安装与初始化.....	错误!未定义书签。
4.2 输入.....	5
4.2.1 输入数据的现实背景.....	5
4.2.2 输入格式.....	6
4.2.3 输入举例.....	6
4.3 输出对每项输出作出说明.....	6
4.3.1 输出数据的现实背景.....	6
4.3.2 输出格式.....	6
4.3.3 输出举例.....	7
4.4 文卷查询.....	8
4.5 出错处理和恢复.....	8
4.6 终端操作.....	8

用户手册（GB8567——88）

1 引言

1.1 编写目的

该用户手册主要用于指导用户更好的使用 Project-Spider 爬虫网站，指导用户更好地应用该系统编写，该文档是为了指导普通用户使用操作该程序。处理和查询想要知道的数据，并为测试提供依据，主要适用于对漫画插画感兴趣的用户对 p 站更好的应用。

1.2 背景

说明：

- a. 软件系统名称：Project-Spider 基于 p 站的爬虫网站
- b. 任务提出者：12 小组
- c. 开发者：12 小组

用户：任何对 p 站漫画、插画感兴趣的用户

1.3 定义

无

1.4 参考资料

参考资料	提交时间	编写人员	备注
《Project-Spider 项目计划书》	2017.3	周林 徐遵杰 游宇杰	无
《Project-Spider 需求规格说明书》	2017.4	周林 徐遵杰 游宇杰	无
《Project-Spider 设计文档》	2017.5	周林 徐遵杰 游宇杰	无
《Project-Spider 测试计划》	2017.5	周林 徐遵杰 游宇杰	无

表 1：引用的项目文件

标题	文件编号	发表日期	出版单位
《计算机软件文档编制规范》	GB/T 8567-2006	2006 年	中国国家标准化管理委员会
《工程化文档规范》	GB/T 26805.4-2011	2011 年 7 月 29 日	中国国家标准化管理委员会
《国家标准软件开发文档模板——用户手册》	GB8567—88	2008 年 4 月 11 日	中国国家标准化管理委员会

表 2：参考的文档编写标准资料

2 用途

2.1 功能

模块	功能点
Index Page	导航、页头及整体布局等
Main Page	主页面设计
Picture Page	插画展示页面、详细展示
Download Page	下载页面、下载功能
Search Page	检索页面、数据分析决策功能
Crawler Page	爬虫功能、获取数据
Crawler Management	爬虫管理
Request Management	Cookie、请求头的获取
File Management	文件模块，记录数据
Data Management	数据分析模块
Database Management	数据库模块
Route Management	路由管理
Test Management	测试模块
Middleware Management	中间件的设计
About Management	主页面后台编辑、管理
News Management	插画分类的后台编辑、管理

Search Management	检索功能后台编辑、管理
Log Management	日志的收集和管理
User Management	收集用户信息
Download Management	下载后台编辑、管理

2.2 性能

2.2.1 精度

无

2.2.2 时间特性

系统相应操作响应时间不超过 3 秒；

用户搜索插画操作从输入检索条件，电脑提交数据到得到查询结果不超过 3 秒；数据管理部分，从提交某一数据录入到结果返回不超过 3 秒。

2.2.3 灵活性

操作方式：支持多种操作系统；

运行环境：支持各种浏览器访问；

用户无需安装客户端、数据库直接通过浏览器访问网站即可。

2.3 安全保密

无

3 运行环境

3.1 硬设备

- 硬件环境
客户机：PC
OS:win10

CPU:2.2Ghz 4 核
内存 4G
服务机: Linux*3
OS: CentOS 6.5 64 位
CPU 2.0Ghz 1 核
内存 1G
带宽: 1Mbps
软件环境
客户机: 浏览器 Edge40,Chrome57,ie11
服务机: 服务器 node
数据库 mongodb

3.2 支持软件

浏览器: 火狐浏览器、chrome 浏览器等

3.3 数据结构

本系统的使用需要连接 mongodb 数据库,当需更新数据库时,可以新建数据库,并且在数据库中按需要新建各个表格,存储在相应的文件夹中。

4 使用过程

4.1 安装与初始化

用户按说明安装浏览器,然后后对打开 project-Spider 网址即可使用系统,查看数据、检索数据,其数据全部保存在服务器上。

4.2 输入

规定输入数据和参量的准备要求。

4.2.1 输入数据的现实背景

说明输入数据的现实背景,主要是

- a. P 站未对普通用户提供检索功能,用户需要检索自己感兴趣的漫画、插画数据。
- b. 用户通过一定条件检索的得到漫画、插画信息及地址。
- c. 每天为一个周期作为爬取数据的更新周期。

- d. 检索条件的输入通过键盘输入。
- e. 用户只具备对漫画插画的查询功能以及检索功能；
- F 对用户输入条件进行记录，用户可以通过看到输入历史。

4.2.2 输入格式

	类型	格式	数值范围	精度
输入	查询	字符串	无	无
	表单查询	JSON 表单	整数	无
输出	图片	JPG/PNG	无	无
	作者/作品信息	字符串	无	无
	错误信息	HTML	无	无

a.

4.2.3 输入举例

为每个完整的输入形式提供样本，包括：

- a. 控制或首部——例如用来表示输入的种类和类型的信息，标识符输入日期，正文起点和对所用编码的规定；
- b. 主体——输入数据的主体，包括数据文卷的输入表述部分；
- c. 尾部——用来表示输入结束的控制信息，累计字符总数等；
- d. 省略——指出哪些输入数据是可省略的；
- e. 重复——指出哪些输入数据是重复的。

4.3 输出对每项输出作出说明

4.3.1 输出数据的现实背景

输出的数据供用户查看和下载，用户可以随时查询，复制地址到迅雷下载插画。

说明输出数据的现实背景，主要是：

- a. 输出的数据供用户查看、下载；
- b. 以每天/周为周期进行信息的爬取、更新；
- c. 为检查信息的正确，可和以往信息对比来确认新输入的信息格式是否相符；

4.3.2 输出格式】

URL, 图片, 压缩包。

4.3.3 输出举例

- a. 当用户点击预览按钮后

```
/**这个页面是抓取到的下载地址，你可以将他们复制到迅雷等下载器进行下载*/  
/**  
共计：5  
*/
```

[/*我想要预览并打包下载*/](#)

```
https://i.pximg.net/img-original/img/2017/03/19/13/09/41/61989393_p0.jpg  
https://i.pximg.net/img-original/img/2017/03/01/06/34/17/61690808_p0.jpg  
https://i.pximg.net/img-original/img/2017/03/20/19/14/52/62013212_p0.jpg  
https://i.pximg.net/img-original/img/2017/06/03/22/44/36/63205494_p0.jpg  
https://i.pximg.net/img-original/img/2017/03/20/23/07/06/62017672_p0.jpg
```

- b. 当点击下载打包下载后

[全部下载](#) 共计 5


点赞数:188
作品ID:61989393_p0_master1200.jpg

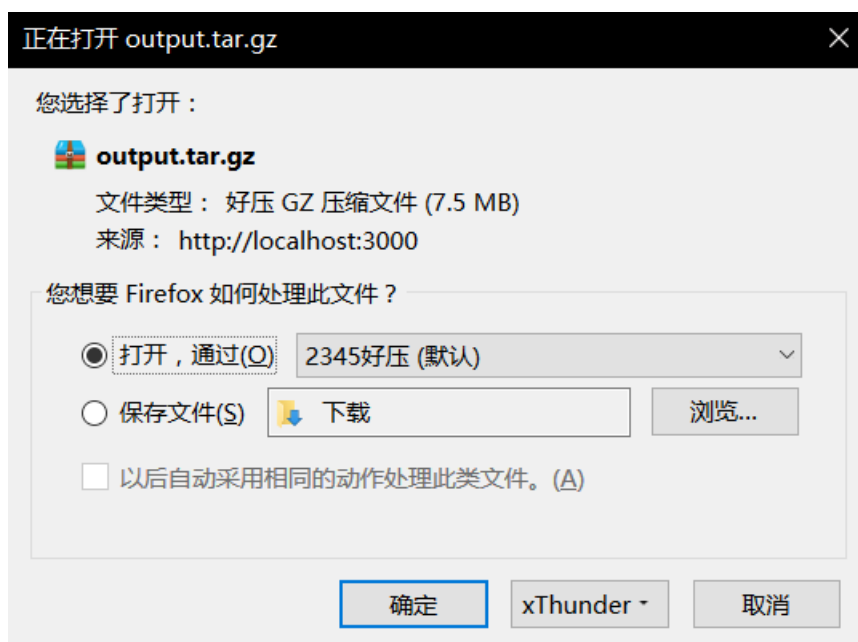

点赞数:148
作品ID:61690808_p0_master1200.jpg


点赞数:16
作品ID:62013212_p0_master1200.jpg


点赞数:16
作品ID:63205494_p0_master1200.jpg


点赞数:4
作品ID:62017672_p0_master1200.jpg

- c. 当点击全部下载后



4.4 文卷查询

用户输入检索条件后，点击“直接查询”或“执行爬虫”按钮即可对漫画、插画信息进行查询

4.5 出错处理和恢复

没有查询结果：当没有查询结果时可能是当前并不存在符合该条件的漫画、插画信息。

4.6 终端操作

无特别说明，用户只需按照按钮的说明操作即可。