基于Scrapy的WebUI开发

——测试需求规格说明书

SY1906514 汪凌风 SY1906508 明　昊

SY1906513 汪丽萍 SY1906431 郑泽西

SY1806220 闫奕涛 SY1906509 邵志钧

2020/05/13

# 版本记录

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 版本 | 日期 | 修改说明 | 参与者 | 审核人 |
| 1.0.0 | 2020.05.13 | 完成第一版 | 全体成员 | 汪凌风 |
| 1.1.0 | 2020.05.18 | 修改部分描述，增加白盒测试内容 | 全体成员 | 闫奕涛 |

注：全体成员即汪凌风，邵志钧，明昊，汪丽萍，闫奕涛，郑泽西

目　录

[一、引言 4](#_Toc40714248)

[1.1 测试目的 4](#_Toc40714249)

[1.2 测试背景 4](#_Toc40714250)

[1.3 测试范围 4](#_Toc40714251)

[1.4 术语与缩略语 4](#_Toc40714252)

[1.5 测试大纲 5](#_Toc40714253)

[1.6 参考资料 5](#_Toc40714254)

[二、测试计划 6](#_Toc40714255)

[2.1 测试方法 6](#_Toc40714256)

[2.2 测试工具 6](#_Toc40714257)

[2.3 测试分工 6](#_Toc40714258)

[2.4 测试环境 7](#_Toc40714259)

[2.5 测试用例对应关系 7](#_Toc40714260)

[三、功能需求测试 9](#_Toc40714261)

[3.1白盒测试过程 9](#_Toc40714262)

[3.1.1 管理脚本模块 9](#_Toc40714263)

[3.1.2 自动化生成脚本模块 10](#_Toc40714264)

[3.2黑盒测试过程 11](#_Toc40714265)

[3.2.1 脚本管理 11](#_Toc40714266)

[3.2.2 脚本生成 13](#_Toc40714267)

[3.2.3 词云生成 14](#_Toc40714268)

[3.2.4 选择优化 14](#_Toc40714269)

[四、非功能需求测试 18](#_Toc40714270)

[4.1 可用性 18](#_Toc40714271)

[4.2 稳定性 18](#_Toc40714272)

[4.3 交互界面用户友好 19](#_Toc40714273)

[4.4 高定制化 19](#_Toc40714274)

# 一、引言

## 1.1 测试目的

本组基于Scapy进行可视化页面开发， 实现了一些关于爬虫的可视化功能。

编写此测试需求规格说明书的目的是指导接下来对于所开发系统的测试活动。测试的目的是检查开发的系统是否完成了软件需求规格说明书中所要求的需求，检查所开发的系统是否正确可靠，发现系统中的问题，最后撰写测试报告。

## 1.2 测试背景

项目名称：基于Scrapy的WebUI开发

测试人员：全体成员

开发人员：全体成员

## 1.3 测试范围

本测试依据软件需求规格说明书与设计文档，对WebUI的功能性需求和非功能性需求进行测试。

对WebUI的功能性需求描述在需求规格说明书的第四章，其测试用例在本文档第三章。对WebUI的非功能性需求描述在需求规格说明书的第四章，其测试用例在本文档第四章。

## 1.4 术语与缩略语

　　Engine，Downloader等组件在下面的组件介绍中有解释。

|  |  |
| --- | --- |
| 爬虫 | 是一种按照一定的规则，自动地抓取[万维网](https://baike.baidu.com/item/%E4%B8%87%E7%BB%B4%E7%BD%91/215515" \t "_blank)信息的程序或者脚本。 |
| Item | 项目，它定义了爬取结果的数据结构，爬取的数据会被赋值成该Item对象。 |
| Selectors | 选择器，Scrapy有自己的数据提取机制，它们被称为选择器。因为它们“选择”HTML文档的某些部分 XPath 或 CSS 表达。 |
| Scraper | 刮取器，主要作用是对Spider中间件进行管理，通过中间件完成请求、响应、数据分析等工作。 |
| Request | 请求，是URL经过Engine组件封装后的，Downloader组件即将接收的内容 |

## 1.5 测试大纲

　　本文共分为四个章节，分别是引言、测试计划、功能需求测试和非功能需求测试。第二章测试计划介绍了项目的测试方法、测试工具以及测试用例对应关系等；第三章功能需求测试分为了白盒测试与黑盒测试两大部分；第四章非功能需求测试介绍了系统级的测试用例。

## 1.6 参考资料

1. 邹欣.构建之法[M].北京.人民邮电出版社,2015:264-296
2. 单元测试[EB/OL].<https://www.liaoxuefeng.com/wiki/897692888725344/953546675792640>
3. 软件工程测试方法[EB/OL]. <https://blog.csdn.net/fightfaith/article/details/47858101>

# 二、测试计划

## 2.1 测试方法

　　对于功能测试，我们选择同时应用黑盒测试与白盒测试的方式进行检查，白盒测试是一种测试用例设计方法，盒子指的是被测试的软件，白盒指的是盒子是可视的，即清楚盒子内部的东西以及里面是如何运作的。"白盒"法全面了解程序内部逻辑结构、对所有逻辑路径进行测试；黑盒测试又称为功能测试，主要检测软件的每一个功能是否能够正常使用。在测试过程中，将程序看成不能打开的黑盒子，不考虑程序内部结构和特性的基础上通过程序接口进行测试，检查程序功能是否按照设计需求以及说明书的规定能够正常打开使用。我们通过结合两者的方式来检查模块程序内部逻辑结构与运行情况和系统功能的正确性与容错性。

## 2.2 测试工具

对于黑盒测试，测试过程的代码撰写和用例测试主要通过Python的unittest包来实现，模块的一些测试结果通过Pycharm的coverage功能进行分析，对于白盒测试，因为本项目将部署在云端服务器上，测试时通过URL即可访问基于Scrapy的WebUI页面，所以选择在页面上人工按照测试需求和用例完成全部测试工作。

## 2.3 测试分工

|  |  |  |
| --- | --- | --- |
| **成员** | **主要职责** | **测试用例** |
| 明昊 | 完成管理脚本功能需求测试，并对WebUI进行总体设计和完善 | 管理脚本 |
| 汪丽萍 | 完成生成脚本功能需求测试，并撰写对应测试文档 | 生成脚本 |
| 郑泽西 | 完成对应功能需求测试，并撰写对应测试文档 | 暂无 |
| 闫奕涛 | 完成生成词云功能需求测试，负责测试需求规格说明书的整体框架，并撰写对应测试文档 | 生成词云 |
| 邵志钧 | 完成选择优化功能需求测试，并撰写对应测试文档 | 选择优化 |
| 汪凌风 | 完成去重优化功能需求测试，并撰写对应测试文档 | 去重优化 |

## 2.4 测试环境

　　Chrome 7.0版本以上，Windows自带edge浏览器等。

## 2.5 测试用例对应关系

|  |  |  |  |
| --- | --- | --- | --- |
| **功能需求** | **需求文档对应章节** | **测试用例** | **测试点** |
| 管理脚本 | 4.3.1 | 运行脚本 | 核对是否能够上传项目 |
| 核对是否能够运行 |
| 查看输出 | 核对是否能够查看道日志输出 |
| 下载输出 | 核对是否能够下载文件 |
| 管理记录 | 核对是否能够删除项目 |
| 管理脚本模块单元测试 | 能否按照测试数据的顺序正确合理地输出预期的结果 |
| 自动化生成脚本 | 4.3.2 | 用例自动生成scrapy项目压缩文件 | 表单提交功能是否正常，scrapy项目生成功能是否正常，文件修改功能是否正常，文件压缩功能是否正常，读取文件功能是否正常 |
| 返回自动生成的scrapy项目压缩文件的文件名 |
| 用例下载scrapy项目压缩文件 | 下载文件功能是否正常，下载文件是否是请求文件 |
| 返回请求的scrapy项目压缩文件 |
| 在线编辑器代码编辑是否正常 |
| 在线调试功能测试 | 代码运行相关信息是否输出正常 |
| 代码运行错误处理是否正常 |
| 自动化生成脚本模块单元测试 | 能否按照测试数据的顺序正确合理地输出预期的结果 |
| 在线运行功能测试 | 正确代码能否正常运行 |
| 运行结果输出与下载是否正常 |
| 生成词云 | 4.3.3 | 选择输入的URL | 核对是否可以正常点击选择框中的URL选项 |
| 生成词云 | 核对是否可以正常显示输入网页对应的词云图片 |
| 选择优化 | 4.3.4 | 随机代理IP测试 | 实时爬取代理IP功能是否正常 |
| 爬虫脚本代码优化是否正常 |
| 爬虫脚本运行是否正常 |
| 爬虫运行是否使用代理IP |
| 随机User-Agent测试 | 爬虫脚本代码优化是否正常 |
| 爬虫脚本运行是否正常 |
| 爬虫运行是否使用随机User-Agent |
| 爬取间隔测试 | 爬虫脚本代码优化是否正常 |
| 爬虫脚本运行是否正常 |
| 爬虫运行页面爬取间隔是否正确 |
| Cookie测试 | 爬虫脚本代码优化是否正常 |
| 爬虫脚本运行是否正常 |
| 爬虫运行是否按要求携带Cookie进行爬取 |
| 去重优化 | 爬虫是否可以选择优化后的去重算法并正确爬取 |

# 三、功能需求测试

按照第二章测试方法和工具的介绍，对于功能需求的测试，我们选择同时应用黑盒测试与白盒测试的方式进行检查，通过Python的unittest包和Pycharm的coverage功能来实现黑盒测试过程的用例测试，通过人工访问基于Scrapy的WebUI页面来完成部分功能的白盒测试。因此在该章节中，项目团队把功能需求的测试分成黑盒和白盒两部分来进行测试需求的制订和测试用例的撰写。

## 3.1白盒测试过程

4对于白盒测试过程，我们通过Python的unittest包和Pycharm的coverage功能来实现部分模块的测试，因为脚本生成和选择优化功能集成在一起，且API何在一起，所以这两个模块的单元测试一起进行。

### 3.1.1 管理脚本模块

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **名称** | 管理脚本模块单元测试 | | | | | |
| **模块功能** | 管理脚本的保存和运行 | | | | | |
| **模块优先级** | 高 | | | | | |
| **测试目的** | 测试API函数合法的输入参数及正确的调用顺序 | | | | | |
| **测试描述** | 在每一个模块测试中，调用所有的API接口函数。对API函数参数输入合法参数值，并且以正确的顺序调用，打印输出函数返回值。 | | | | | |
| **前提和约束** | Redis与Scrapyd正常运行 | | | | | |
| **测试用例** | **序号** | **输入数据** | **预期输出** | **实际输出** | **是否通过** | **备注** |
| 1 | File: ../DistributedSpider-master.zip,  Project: DistributedSpiders,  Spider: Dis\_Douban | success |  |  |  |
| 2 | Name: DistributedSpiders, | success |  |  |  |
| 3 | Name: DistributedSpiders, | such project is running or pending |  |  |  |
| 4 | Name：dis | no such project exists |  |  |  |
| 5 | Name: DistributedSpiders | Fail,  such project is running or pending |  |  |  |
| 6 | Name: DistributedSpiders | Success |  |  |  |
| 7 | Name：etstdemo | such project not exist |  |  |  |
| 8 | Name: DistributedSpiders | Success |  |  |  |
| 9 | 无 | False |  |  |  |
| **覆盖率** |  | | | | | |
| **备注** |  | | | | | |

### 3.1.2 自动化生成脚本模块

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **名称** | 自动化生成脚本模块单元测试 | | | | | |
| **模块功能** | 自动生成爬虫脚本并压缩项目 | | | | | |
| **模块优先级** | 无 | | | | | |
| **测试目的** | 测试API函数合法的输入参数及正确的调用顺序 | | | | | |
| **测试描述** | 在每一个模块测试中，调用所有的API接口函数。对API函数参数输入合法参数值，并且以正确的顺序调用，打印输出函数返回值。 | | | | | |
| **前提和约束** | 无 | | | | | |
| **测试用例** | **序号** | **输入数据** | **预期输出** | **实际输出** | **是否通过** | **备注** |
| 1 | test.json文件 | success |  |  |  |
| 2 | 无 | “name”:”demo.zip” |  |  |  |
| **覆盖率** |  | | | | | |
| **备注** |  | | | | | |

## 3.2黑盒测试过程

对于黑盒测试过程，我们通过人工访问基于Scrapy的WebUI页面来完成部分功能的测试，按照不同模块的功能与需求的对应关系，我们按照2.5的表格制定了一下测试用例。

### 3.2.1 脚本管理

|  |  |  |  |
| --- | --- | --- | --- |
| Test Case Specification | | | |
| **名称** | 运行脚本 | | |
| **简要描述** | 本测试验证用户是否能运行脚本 | | |
| **前提和约束** | 用户打开WebUI | | |
| **测试步骤** | **1** | 用户点击选择脚本上传项目与必要参数 | |
| **2** | 用户选中项目点击运行 | |
| **预期结果** | 脚本状态变为运行 | |
| **评价准则** | 项目的状态是否变为运行 | | |
| **异常情况** | 项目启动失败 | | 提示“项目启动失败” |

|  |  |  |  |
| --- | --- | --- | --- |
| Test Case Specification | | | |
| **名称** | 查看输出 | | |
| **简要描述** | 本测试验证用户是否能查看日志输出 | | |
| **前提和约束** | 用户打开WebUI | | |
| **测试步骤** | **1** | 用户选择脚本 | |
| **2** | 用户点击查看输出 | |
| **3** | 界面弹出小框显示输出 | |
| **预期结果** | 在弹出小框内正常显示日志输出 | |
| **评价准则** | 用户是否可以看到日志输出运行脚本 | | |
| **异常情况** | 该项目没有保存的日志输出 | | 提示“输出文件不存在” |

|  |  |  |  |
| --- | --- | --- | --- |
| Test Case Specification | | | |
| **名称** | 下载输出 | | |
| **简要描述** | 本测试验证用户是否能下载输出 | | |
| **前提和约束** | 用户选择脚本查看输出 | | |
| **测试步骤** | **1** | 用户点击下载输出按钮 | |
| **预期结果** | 浏览器开始下载文件 | |
| **评价准则** | 下载文件内容是否与看到的输出结果一致 | | |
| **异常情况** | 下载文件不存在 | | 提示“下载文件不存在” |

|  |  |  |  |
| --- | --- | --- | --- |
| Test Case Specification | | | |
| **名称** | 管理记录 | | |
| **简要描述** | 本测试验证系统是否能够保存用户的过往爬取信息 | | |
| **前提和约束** | 用户在WebUI内运行过项目 | | |
| **测试步骤** | **1** | 用户重新打开WebUI | |
| **预期结果** | 用户可以看到自己之前上传的项目并且查看输出 | |
| **评价准则** | 是否可以看到之前上传的项目 | | |
| **异常情况** | 看不到之前的项目 | | 无提示 |

### 3.2.2 脚本生成

|  |  |  |  |
| --- | --- | --- | --- |
| Test Case Specification | | | |
| **名称** | 自动生成scrapy项目压缩文件测试 | | |
| **简要描述** | 本测试验证用户是否能通过提交表单相应信息自动生成scrapy项目压缩文件 | | |
| **前提和约束** | 用户运行系统 | | |
| **测试步骤** | **1** | 选择“自动生成脚本”子功能 | |
| **2** | 填写项目名，网址，item和对应xpath等表单内容 | |
| **3** | 选择“提交”按钮进行提交 | |
| **预期结果** | 表格显示自动生成的scrapy项目压缩文件的文件名 | |
| **评价准则** | 实际结果与预期结果一致 | | |
| **异常情况** | 表单未正确填写 | | 提示相应的表单校验规则信息 |

|  |  |  |
| --- | --- | --- |
| Test Case Specification | | |
| **名称** | 下载scrapy项目压缩文件测试 | |
| **简要描述** | 本测试验证用户是否能下载自动生成的scrapy项目压缩文件 | |
| **前提和约束** | 用户自动生成的scrapy项目压缩文件名显示在表格上 | |
| **测试步骤** | **1** | 点击需要下载的scrapy项目压缩文件的“下载”按钮 |
| **预期结果** | 浏览器下载相应scrapy项目压缩文件 |
| **评价准则** | 实际结果与预期结果一致 | |
| **异常情况** | 无 | |

### 3.2.3 词云生成

|  |  |  |  |
| --- | --- | --- | --- |
| Test Case Specification | | | |
| **名称** | 生成词云 | | |
| **简要描述** | 用户在选择框中选取URL，该用例能够成功生成词云图 | | |
| **前提和约束** | 用户进入生成词云页面 | | |
| **测试步骤** | **1** | 在选择框中选取URL | |
| **2** | 点击“生成词云” | |
| **3** | 运行脚本 | |
| **预期结果** | 脚本正确运行并成功生成词云 | |
| **评价准则** | 词云生成图片正确且完整，运行结果与预期结果一致 | | |
| **异常情况** | 爬取对应URL失败 | | 提示“爬取失败” |
| 脚本生成错误 | | 运行提示脚本异常 |

### 3.2.4 选择优化

|  |  |  |  |
| --- | --- | --- | --- |
| Test Case Specification | | | |
| **名称** | 随机代理IP测试 | | |
| **简要描述** | 本测试验证用户选择随机代理IP优化选项后，能否生成正确的脚本并正确运行 | | |
| **前提和约束** | 用户进入自动化生成脚本页面 | | |
| **测试步骤** | **1** | 选择“随机代理IP”优化选项并点击生成脚本 | |
| **2** | 检查脚本内容 | |
| **3** | 运行脚本 | |
| **预期结果** | 脚本正确运行并使用代理IP进行爬取 | |
| **评价准则** | 生成脚本内容正确且完整，运行结果与预期结果一致 | | |
| **异常情况** | 脚本生成错误 | | 运行提示脚本异常 |

|  |  |  |  |
| --- | --- | --- | --- |
| Test Case Specification | | | |
| **名称** | 随机User-Agent测试 | | |
| **简要描述** | 本测试验证用户选择随机User-Agent优化选项后，能否生成正确的脚本并正确运行 | | |
| **前提和约束** | 用户进入自动化生成脚本页面 | | |
| **测试步骤** | **1** | 选择“随机User-Agent”优化选项并点击生成脚本 | |
| **2** | 检查脚本内容 | |
| **3** | 运行脚本 | |
| **预期结果** | 脚本正确运行并使用随机User-Agent进行爬取 | |
| **评价准则** | 生成脚本内容正确且完整，运行结果与预期结果一致 | | |
| **异常情况** | 脚本生成错误 | | 运行提示脚本异常 |

|  |  |  |  |
| --- | --- | --- | --- |
| Test Case Specification | | | |
| **名称** | 爬取间隔测试 | | |
| **简要描述** | 本测试验证用户设置爬取间隔后，能否生成正确的脚本并正确运行 | | |
| **前提和约束** | 用户进入自动化生成脚本页面 | | |
| **测试步骤** | **1** | 选择随机代理IP后设置优化选项“页面爬取间隔”参数并点击生成脚本 | |
| **2** | 检查脚本内容 | |
| **3** | 运行脚本 | |
| **预期结果** | 脚本正确运行并使用设定页面爬取间隔进行爬取 | |
| **评价准则** | 生成脚本内容正确且完整，运行结果与预期结果一致 | | |
| **异常情况** | 脚本生成错误 | | 运行提示脚本异常 |

|  |  |  |  |
| --- | --- | --- | --- |
| Test Case Specification | | | |
| **名称** | Cookie测试 | | |
| **简要描述** | 本测试验证用户设置是否携带Cookie后，能否生成正确的脚本并正确运行 | | |
| **前提和约束** | 用户进入自动化生成脚本页面 | | |
| **测试步骤** | **1** | 设置优化选项“是否携带Cookie”并点击生成脚本 | |
| **2** | 检查脚本内容 | |
| **3** | 运行脚本 | |
| **预期结果** | 脚本正确运行并按要求携带Cookie进行爬取 | |
| **评价准则** | 生成脚本内容正确且完整，运行结果与预期结果一致 | | |
| **异常情况** | 脚本生成错误 | | 脚本生成错误 |

|  |  |  |  |
| --- | --- | --- | --- |
| Test Case Specification | | | |
| **名称** | 选择去重算法优化 | | |
| **简要描述** | 本测试验证用户选择去重优化后能否正常运行以及内存消耗是否降低 | | |
| **前提和约束** | 用户运行一个能够正确运行的爬虫项目 | | |
| **测试步骤** | **1** | 点击选择去重算法优化 | |
| **2** | 点击确定运行优化后的项目 | |
| **预期结果** | 爬虫项目与为选择优化时最终结果一致，且在百万级URL数目的项目中，服务器内存占用比原先降低。 | |
| **评价准则** | 实际结果与预期结果一致 | | |
| **异常情况** | URL去重出现误判 | | URL数目过多（亿级），修改配置文件中的参数，使算法能够适应URL数目 |

# 四、非功能需求测试

## 4.1 可用性

|  |  |  |  |
| --- | --- | --- | --- |
| Test Case Specification | | | |
| **名称** | 可用性测试 | | |
| **简要描述** | 本测试验证应用运行时，管理脚本、自动化生成脚本、生成词云、在线编辑调试和选择优化四大主要功能是否能正常运行 | | |
| **前提和约束** | 用户打开ScrapyWebUI网址 | | |
| **测试步骤** | **1** | 用户进行各项功能的使用 | |
| **2** | 测试各项功能是否正常运行，是否存在功能失效的情况 | |
| **预期结果** | 各项功能正常运行，不存在功能失效的情况 | |
| **评价准则** | 实际结果与预期结果一致 | | |
| **异常情况** | 无 | |  |

## 4.2 稳定性

|  |  |  |  |
| --- | --- | --- | --- |
| Test Case Specification | | | |
| **名称** | 稳定性测试 | | |
| **简要描述** | 本测试验证从用户打开应用开始，2小时以内可以稳定运行无故障，出故障时，能够发出故障信息 | | |
| **前提和约束** | 用户打开ScrapyWebUI网址 | | |
| **测试步骤** | **1** | 用户进行各项功能的使用 | |
| **2** | 测试从用户打开网址开始，是否可以2小时以内稳定运行无故障，若出现故障，是否能够发出故障信息 | |
| **预期结果** | 从用户打开网址开始，2小时以内可以稳定运行无故障，出现故障时，能够发出故障信息 | |
| **评价准则** | 实际结果与预期结果一致 | | |
| **异常情况** | 无 | |  |

## 4.3 交互界面用户友好

|  |  |  |  |
| --- | --- | --- | --- |
| Test Case Specification | | | |
| **名称** | 交互界面用户友好性测试 | | |
| **简要描述** | 本测试验证应用运行时，其实用过程是否有较为全面的用户使用说明、提示及引导 | | |
| **前提和约束** | 用户启动本应用 | | |
| **测试步骤** | **1** | 用户启动应用 | |
| **2** | 用户进行各项功能的使用 | |
| **3** | 检查各项功能的用户引导是否完整 | |
| **预期结果** | 用户提示及引导完善且正常显示 | |
| **评价准则** | 实际结果与预期结果一致 | | |
| **异常情况** | 无 | |  |

## 4.4 高定制化

|  |  |  |  |
| --- | --- | --- | --- |
| Test Case Specification | | | |
| **名称** | 高定制化测试 | | |
| **简要描述** | 本测试验证用户可以使用本系统是否可以实现高自由度的爬虫定制 | | |
| **前提和约束** | 用户进入自动化生成脚本页面 | | |
| **测试步骤** | **1** | 用户进入自动化生成脚本页面 | |
| **2** | 用户对爬虫进行定制 | |
| **3** | 用户使用定制爬虫可以爬取大部分不使用动态JS编写的网站 | |
| **预期结果** | 爬虫脚本运行正常。 | |
| **评价准则** | 实际结果与预期结果一致 | | |
| **异常情况** | 无 | |  |