|  |
| --- |
|  |
| 《大数据事件智能抓取和画像系统》测试报告 |
| **参选赛题：基于互联网大数据的事件智能抓取和画像** |
|  |
| **参赛者：霍震霖 余泽雄 林盛环** |

**指导老师：花罡辰**

|  |
| --- |
| **参赛单位：深圳信息职业技术学院软件学院 2017年6月3日** |

目录

[第一章 引言 1](#_Toc486526009)

[1.1编写目的 1](#_Toc486526010)

[1.2参考文档 1](#_Toc486526011)

[第二章 测试用例及测试结果 2](#_Toc486526012)

[2.1 测试用例设计 2](#_Toc486526013)

## 第一章 引言

### 1.1编写目的

测试用例和测试报告文档主要在于测试《大数据事件智能抓取和画像系统》的功能和非功能是否符合需求要求，从中查找出软件缺陷并进行修复。

### 1.2参考文档

《大数据事件智能抓取和画像系统》软件设计文档。

## 第二章 测试用例及测试结果

### 2.1 测试用例设计

#### 2.1.1 测试方法和测试环境

采用单元测试的方法进行测试，主要对核心功能进行测试。用Python单元测试框架pyunit进行验证。

测试运行的软硬件环境如下：

|  |  |
| --- | --- |
| CPU | Intel酷睿 i5 4590及以上 |
| 内存 | 4GB及以上 |
| 硬盘 | 50GB及以上 |
| 操作系统 | Microsoft Windows 10, 64位操作系统 |
| 其它软件 | CPython 2.7.10以上，推荐2.7.13  XAMPP 3.2.2，MySQL 5.7  Chrome 59.0.3071.109以上 |
| 主要开发库版本 | beautifulsoup 4.6.0，pandas 0.20.2，requests 2.18.1，selenium 3.4.3，pyltp 0.1.9.1，newspaper 0.1.0.7 |

表 2-1 测试运行环境

#### 2.1.2 测试用例

* 政治会议属性提取

|  |  |  |
| --- | --- | --- |
| **用例编号** | | DG-1-1 |
| **测试内容** | softbei/void/huiyi\_nlp.py **get\_tree**(words, arcs, postags, root\_string) | |
| **前置条件** | 软件能够正常访问互联网、语料库配置正常 | |
| **序号** | | **输入数据** |
|  | | 输入参数依次为以下文字的（分词结果、语法树分析结果、词性分析结果、字符串“举行”）， "2017年夏季达沃斯论坛新领军者年会将于6月27日至29日在大连举行，目前共有来自80多个国家的超过2000位政、商、学、文等各界领袖注册参会。本次会议主题为“在第四次工业革命中实现包容性增长”。 |
| **预期结果** | 提取出的会议名、举办时间，举办地点分别为：2017年夏季达沃斯论坛新领军者年会、6月27日 29日、大连 | |
| **测试结果** | **通过** | |

|  |  |  |
| --- | --- | --- |
| **用例编号** | | DG-1-2 |
| **测试内容** | softbei/void/huiyi\_nlp.py **get\_tree**(words, arcs, postags, root\_string) | |
| **前置条件** | 软件能够正常访问互联网、语料库配置正常 | |
| **序号** | | **输入数据** |
|  | | 输入参数依次为以下文字的（分词结果、语法树分析结果、词性分析结果、字符串“举行”）， "中青在线杭州6月22日电（实习生 管婷婷 中国青年报·中青在线记者 董碧水）记者从杭州市人民政府今天举行的新闻发布会获悉，2017全球私募基金西湖峰会将于6月24日在杭州召开。”  。 |
| **预期结果** | 提取出的会议名、举办时间，举办地点分别为：2017 全球 私募基金西湖峰会、6月 24日、杭州 | |
| **测试结果** | **通过** | |

* 事件去重

|  |  |  |
| --- | --- | --- |
| **用例编号** | | DG-2-1 |
| **测试内容** | softbei/csv\_removal/class\_removal.py **drop\_repet**(csv\_name, colunm, update\_name) | |
| **前置条件** | 软件能够正常访问互联网 | |
| **序号** | | **输入数据** |
|  | | 注：drop\_repet()方法接收的参数分别为待去重的csv文件名，列名，和去重后数据存放的csv文件名，这里从列中抽取两个测试用例进行测试  [万有音乐系]恋\*咫尺回忆—“岛歌王子”中孝介2017巡回演唱会 北京站，  [万有音乐系]恋\*咫尺回忆—“岛歌王子”中孝介2017巡回演唱会 |
| **预期结果** | 两件事情为同一件事情，去除第二条重复数据 | |
| **测试结果** | **通过** | |

|  |  |  |
| --- | --- | --- |
| **用例编号** | | DG-2-2 |
| **测试内容** | softbei/csv\_removal/class\_removal.py **drop\_repet**(csv\_name, colunm, update\_name) | |
| **前置条件** | 软件能够正常访问互联网 | |
| **序号** | | **输入数据** |
| **2** | | 注：drop\_repet()方法接收的参数分别为待去重的csv文件名，列名，和去重后数据存放的csv文件名，这里从列中抽取两个测试用例进行测试  周杰伦2017“地表最强”世界巡回演唱会—重庆站，  “地表最强”2017周杰伦世界巡回演唱会 天津站 |
| **预期结果** | 两件事情为不同事情，没有执行去除第二条步骤 | |
| **测试结果** | **通过** | |

* 事件热度

|  |  |  |
| --- | --- | --- |
| **用例编号** | | DG-3-1 |
| **测试内容** | Softbei/analyse\_data/hot/class\_hot\_num.py get\_hot(title) | |
| **前置条件** | 软件能够正常访问互联网 | |
| **序号** | | **输入数据** |
| **1** | | 周杰伦2017“地表最强”世界巡回演唱会—重庆站 |
| **预期结果** | 热度等级为3 | |
| **测试结果** | **通过** | |

|  |  |  |
| --- | --- | --- |
| **用例编号** | | DG-3-2 |
| **测试内容** | Softbei/analyse\_data/hot/class\_hot\_num.py get\_hot(title) | |
| **前置条件** | 软件能够正常访问互联网 | |
| **序号** | | **输入数据** |
| **2** | | 2017第19届上海国际别墅配套设施博览会 |
| **预期结果** | 热度等级为1 | |
| **测试结果** | **通过** | |

* 关键词提取

|  |  |  |
| --- | --- | --- |
| **用例编号** | | DG-4-1 |
| **测试内容** | Softbei/analyse\_data/age/tf\_idf.py get\_num(title): | |
| **前置条件** | 软件能够正常访问互联网，分词模块加载正常 | |
| **序号** | | **输入数据** |
| **1** | | 2017[A CLASSIC TOUR学友.经典] 世界巡回演唱会—西安站 |
| **预期结果** | 提取出来的关键字为：A CLASSIC TOUR学友.经典，演唱会 | |
| **测试结果** | **通过** | |

|  |  |  |
| --- | --- | --- |
| **用例编号** | | DG-4-2 |
| **测试内容** | Softbei/analyse\_data/age/tf\_idf.py get\_num(title): | |
| **前置条件** | 软件能够正常访问互联网，分词模块加载正常 | |
| **序号** | | **输入数据** |
| **2** | | 五月天LIFE《人生无限公司》2017 MAYDAY LIFE TOUR-北京站 |
| **预期结果** | 提取出来的关键字为：人生无限公司，MAYDAY LIFE TOUR，LIFE | |
| **测试结果** | **通过** | |

* 影响人群

|  |  |  |
| --- | --- | --- |
| **用例编号** | | DG-5-1 |
| **测试内容** | Softbei/analyse\_data/age/class\_getage.py | |
| **前置条件** | 软件能够正常访问互联网，分词模块加载正常 | |
| **序号** | | **输入数据** |
| **1** | | 2017周杰伦世界巡回演唱会厦门站（10月22日场次） |
| **预期结果** | 主要影响年龄层为成年 | |
| **测试结果** | **通过** | |

|  |  |  |
| --- | --- | --- |
| **用例编号** | | DG-5-2 |
| **测试内容** | Softbei/analyse\_data/age/class\_getage.py | |
| **前置条件** | 软件能够正常访问互联网，分词模块加载正常 | |
| **序号** | | **输入数据** |
| **2** | | “梁静茹·你的名字是爱情”2017世界巡回演唱会深圳站（最终场） |
| **预期结果** | 主要影响年龄层为成年 | |
| **测试结果** | **通过** | |

* 主办方级别提取

|  |  |  |
| --- | --- | --- |
| **用例编号** | | DG-6-1 |
| **测试内容** | Softbei/analyse\_data/zhuban/class\_getzhuban.py | |
| **前置条件** | 软件能够正常访问互联网，分词模块加载正常 | |
| **序号** | | **输入数据** |
| **1** | | 中国对外贸易广州展览总公司 |
| **预期结果** | 是国内行业协会 | |
| **测试结果** | **通过** | |

|  |  |  |
| --- | --- | --- |
| **用例编号** | | DG-6-2 |
| **测试内容** | Softbei/analyse\_data/zhuban/class\_getzhuban.py | |
| **前置条件** | 软件能够正常访问互联网，分词模块加载正常 | |
| **序号** | | **输入数据** |
| **2** | | 深圳市宝安区人民政府 |
| **预期结果** | 是地方级政府 | |
| **测试结果** | **通过** | |

* 历史悠久程度

|  |  |  |
| --- | --- | --- |
| **用例编号** | | DG-7-1 |
| **测试内容** | Softbei/analyse\_data/history/class\_history.py get\_value(),get\_num(num) | |
| **前置条件** | 软件能够正常访问互联网，分词模块加载正常 | |
| **序号** | | **输入数据** |
| **1** | | 注：get\_value()方法会调用get\_num(num)方法,get\_num(num)中的参数num为分词后得到的届数  2017第十七届中国（深圳）国际品牌服装服饰交易会 |
| **预期结果** | 历史悠久程度为3 | |
| **测试结果** | **通过** | |

|  |  |  |
| --- | --- | --- |
| **用例编号** | | DG-7-2 |
| **测试内容** | Softbei/analyse\_data/history/class\_history.py get\_value(),get\_num(num) | |
| **前置条件** | 软件能够正常访问互联网，分词模块加载正常 | |
| **序号** | | **输入数据** |
| **2** | | 注：get\_value()方法会调用get\_num(num)方法,get\_num(num)中的参数num为分词后得到的届数  2017中国特许加盟展 |
| **预期结果** | 历史悠久程度为1 | |
| **测试结果** | **通过** | |

* 场馆城市名提取、城市经纬度提取

|  |  |  |
| --- | --- | --- |
| **用例编号** | | DG-8-1 |
| **测试内容** | Softbei/analyse\_data/place/class\_getcity.py to\_place\_csv(csv\_name) | |
| **前置条件** | 软件能够正常访问互联网，ak密匙有效 | |
| **序号** | | **输入数据** |
| **1** | | 注：to\_place\_csv(csv\_name)中的csv\_name为从指定csv文件名字，方法会从csv中读取存放场馆的列，这里选择列中值为“广西体育馆”作为测试用例  广西体育馆 |
| **预期结果** | Lng:108.32297969604537 lat:22.81079813143598  city: 南宁市 | |
| **测试结果** | **通过** | |

|  |  |  |
| --- | --- | --- |
| **用例编号** | | DG-8-2 |
| **测试内容** | Softbei/analyse\_data/place/class\_getcity.py getlal(lal),to\_place\_csv(csv\_name) | |
| **前置条件** | 软件能够正常访问互联网，ak密匙有效 | |
| **序号** | | **输入数据** |
| **2** | | 注：to\_place\_csv(csv\_name)中的csv\_name为从指定csv文件名字，方法会从csv中读取存放场馆的列，这里选择列中值为“国家奥林匹克中心体育场”作为测试用例  国家奥林匹克中心体育场 |
| **预期结果** | Lng:117.17985678210212 lat:39.080385842532664  City: 北京市 | |
| **测试结果** | **通过** | |

* 可视化技术

|  |  |  |
| --- | --- | --- |
| **用例编号** | | DG-9-1 |
| **测试内容** | Softbei/test/page.php页面 | |
| **前置条件** | 服务器正常服务，软件能够正常访问互联网，echarts库加载正常 | |
| **序号** | | **输入数据** |
| **1** | | http://123.207.88.152/softbei/test/page.php?key=2017/06/28,2017/6/28 |
| **预期结果** | 2017年6月28号的数据全部显示正常 | |
| **测试结果** | **通过** | |

|  |  |  |
| --- | --- | --- |
| **用例编号** | | DG-9-2 |
| **测试内容** | Softbei/test/page.php页面，Softbei/test/index.php页面 | |
| **前置条件** | 服务器正常服务，软件能够正常访问互联网，echarts库加载正常 | |
| **序号** | | **输入数据** |
| **2** | | http://123.207.88.152/softbei/test/page.php |
| **预期结果** | 不带参数进行get请求，重定向回<http://123.207.88.152/softbei/test/>index.php页面 | |
| **测试结果** | **通过** | |

### 2.2 性能测试结果

为了验证程序性能，我们抓取了100条事件并提取属性，计算软件花费的时间，并记录内存消耗情况。测试环境如表2-1所示，测试结果如下表所示：

|  |  |
| --- | --- |
| 抓取事件（包含属性提取）件数 | 100 |
| 花费时间 | 约40分钟 |
| 内存消耗峰值 | 1.5GB |

表 2-2 性能测试结果