# <TVMovieJudger >开发文档

版本修订日志

|  |  |  |
| --- | --- | --- |
| 修订时间 | 修订人 | 修订内容 |
| 2016年10月20日 | 杨恬 | 创建 |
|  |  |  |
|  |  |  |
|  |  |  |

## 1．1 TVMovieJudger 豆瓣区分电影电视剧程序

1. 程序编号：TVMovieJudger
2. 程序名称：豆瓣区分电影电视剧程序
3. 程序说明：本程序为根据豆瓣数据区分电影电视剧程序，目前是为精确定位词语含义功能研发，主要研发思路：用豆瓣搜索该名称，处理结果页的第一页，如果有名称相同的就点进去，根据其是否分集判断其为电视剧还是电影，或者两者皆是，准确率可达90%以上，包含的程序模块如下：

模块A：电影电视剧区分功能

1. 程序功能可视化说明：

程序首先读入一个名称，如：致青春。

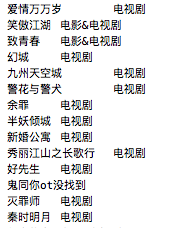
访问豆瓣搜索页：



选出名字（包括别名里有致青春）的网页，点击进去。根据有无集数这个指标判断其类型：



如上是一个名为致青春的电视剧。对每一个名称做如上操作从而得到如下格式的输出文件：



1. 程序输入与输出说明：

输入：电视剧or电影名称列表

输出: 名称 类型

1. 处理过程流程图：

../../../judge.pdf

1. 开放接口（若被调用，需要提供开放的接口说明）：

无

1. 引入的包及接口：

scrapy：运用其中的Selector类进行html解析和元素定位

Levenshtein：判断两个词语的相似度

1. 程序源码目录结构说明：

.

├── \_\_configure\_\_

│   ├── ConfigureParser.py

│   ├── \_\_init\_\_.py

│   └── configure.xml

├── \_\_doc\_\_

│   ├── TVMovieJudger文档.docx

├── \_\_file\_\_

│   ├── baidu\_video.txt

│   └── finished.txt

├── \_\_spider\_\_

│   ├── \_\_init\_\_.py

│   └── doubanJudge.py

├── \_\_util\_\_

│   ├── \_\_init\_\_.py

│   └── library.py

├── main.py

└── run.sh

\_\_configure\_\_: 包含配置文件和配置文件解析类

\_\_doc\_\_：包含程序文档

\_\_file\_\_：包含输入输出以及临时文件

\_\_spider\_\_: 包含爬虫代码

\_\_util\_\_：包含一些常用的函数

main.py: 启动程序

run.sh：执行文件

1. 程序启动方式说明

修改run.sh执行权限，执行run.sh

1. 程序配置文件及说明

配置文件: configure.xml

input\_file: 输入文件

output\_file: 输出文件

1. 部署程序打包方式：

上传文件夹

1. 程序部署说明：

暂无

1. 程序潜在的问题，推荐解决方案及其他特殊说明

豆瓣容易被封，建议停顿3-5秒，或者半夜运行（貌似不需要停顿）

1. 版本号和CHANGES

TVMovieJudger 1.0