**scrapyjs使用文档**

1：什么是scrapyjs？

scrapyjs是由scrapy开发团队开发出的基于splash的解析js的框架

源码地址：https://github.com/scrapinghub/scrapyjs

2：scrapyjs的工作机制是什么？

scrapyjs是利用splash框架来解析js，其作用就是将请求的url发送到splash服务器，由splash来解析js

3：什么是splash？

splash是Lightweight, scriptable browser as a service with an HTTP API

源码地址：https://github.com/scrapinghub/splash

4：环境搭建

a、安装docker

根据官方文档安装https://docs.docker.com/installation/ubuntulinux/

注意：Ubuntu必须为64bit，否则不能正常安装

满足前提条件即可安装：

wget -qO- https:*//get.docker.com/ | sh*

b、通过docker获取scrpinghub/splash项目，并启动服务监听一个端口

docker run -p 8050:8050 scrapinghub/splash

c、安装scrapy&scrapyjs

sudo apt-get install python-pip python-dev build-essential

pip install scrapy

pip install scrapyjs

d、配置scrapy的settings.py

SPLASH\_URL = 'http://192.168.59.103:8050' #splash的服务地址，即b步启动的服务

DOWNLOADER\_MIDDLEWARES = {'scrapyjs.SplashMiddleware': 725,} # Order 725 is just before HttpProxyMiddleware (750) in default scrapy settings. 设置scrapy的downloader

DUPEFILTER\_CLASS = 'scrapyjs.SplashAwareDupeFilter'#虑重过滤器

HTTPCACHE\_STORAGE = 'scrapyjs.SplashAwareFSCacheStorage'#http cache

e、实例