usage

1.运行容器

1. 调出命令终端：

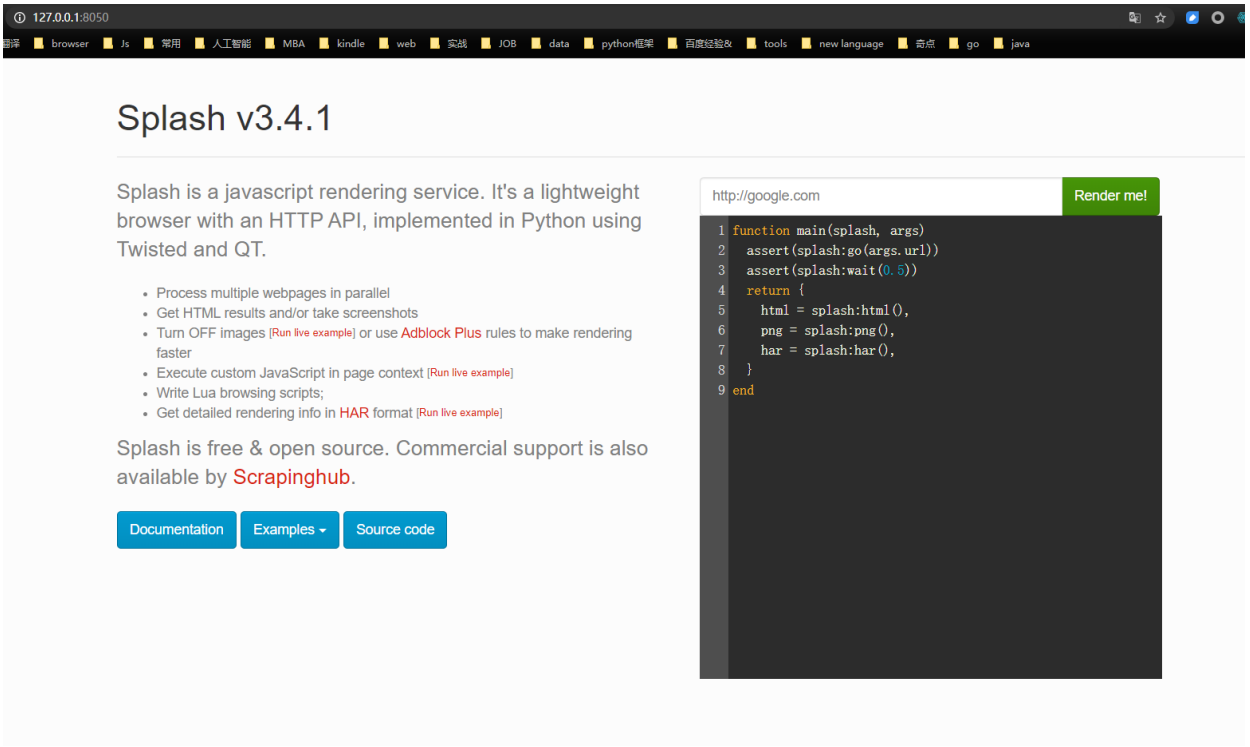
1.1win + r

1.2 输入cmd ,按回车

2. 启动splash服务，如果第一次启动，会从官方拉splash镜像，耗时较长

```
scrapy-splash，这个容器是用来模拟浏览器的请求解析页面：
docker run -it -p 8050:8050 --rm scrapinghub/splash
```

1. 验证：在浏览器中输入：http://127.0.0.1:8050/



2.爬取命令行示例

比如爬取杭州的java职位信息：

在运行命令时需要切换到项目内部中：

```
(zhiping) D:\Project File\Projects\zhipin-spider-master>dir
 驱动器 D 中的卷是 work
 卷的序列号是 DA18-EBFA

 D:\Project File\Projects\zhipin-spider-master 的目录

2020/05/11  16:44    <DIR>          .
2020/05/11  16:44    <DIR>          ..
2020/05/28  17:15    <DIR>          .idea
2020/05/11  16:58    <DIR>          .scrapy
2020/04/23  23:00    <DIR>          .vscode
2020/05/28  16:55    <DIR>          boss_spider
2020/04/23  23:00               575 README.MD
2020/04/23  23:00               265 scrapy.cfg
               2 个文件            840 字节
               6 个目录 324,957,163,520 可用字节

(zhiping) D:\Project File\Projects\zhipin-spider-master>
```

```
scrapy crawl bossspider -a location='hangzhou' -a position='java'
```

如果想要输出到文件:

-a 表示传入参数

-o 表示向外输出

```
scrapy crawl bossspider -a location='hangzhou' -a position='java' -o data.csv
```

注意事项

注意修改settings.py文件里面的mongoDB配置。