

# 使用Scrapy爬虫框架

对百度搜索和百度旗下相关搜索进行关键词爬取

爬取的主要内容包括：标题 时间 平台 作者 时间 简介 正文 链接

使用的时候注意要修改关键词和起始爬虫的URL

## 爬虫使用方式

---

### 单个关键词

#

```
scrapy crawl searchspider -a search=xxxx -a keyword=xxxx
```

或者

```
scrapy crawl newsspider -a search=xxxx -a keyword=xxxx
```

### 使用csv文件

#

在下列路径中放置需要使用的csv文件

BaiduSearch\_Spider\keyword

然后直接使用下列命令：

```
scrapy crawl searchspider
```

## 网页界面启动方式

---

需要安装 flask 框架

```
python Server/main.py
```