

scrapy

Web-crawling framework.

- Create a project:

```
scrapy startproject {{project_name}}
```

- Create a spider (in project directory):

```
scrapy genspider {{spider_name}} {{website_domain}}
```

- Edit spider (in project directory):

```
scrapy edit {{spider_name}}
```

- Run spider (in project directory):

```
scrapy crawl {{spider_name}}
```

- Fetch a webpage as scrapy sees it and print source in stdout:

```
scrapy fetch {{url}}
```

- Open a webpage in the default browser as scrapy sees it (disable javascript for extra fidelity):

```
scrapy view {{url}}
```

- Open scrapy shell for url, which allows interaction with the page source in python shell (or ipython if available):

```
scrapy shell {{url}}
```