

# Scrapyd 配置文件详解

Scrapyd 的配置文件是非常重要的，配置文件中除了日志目录、日志数量、并发数等基础设置外还包括端口设置、IP 绑定设置、API 路由配置等应用层面的设置。

我们需要对 Scrapyd 的配置文件有清晰的认知，避免在认知模糊的情况下对其配置文件进行改动。而且在后续的小节中，也会介绍到如何在代码中读取配置文件以及对应的配置参数值。

## 配置文件参数详解

从之前的小节中我们可以看到，Scrapyd 会在几个核心类中读取 Scrapyd 的配置文件，获取相应的配置参数。那么，配置文件存放在哪里呢？

待你在 Python 环境下安装好 Scrapyd 后，它会在 Python 安装路径下的 `site-packages` 目录下的 `scrapyd` 目录内，配置文件名为 `default_scrapyd.conf`，其内容如下：

```
1  [scrapyd]
2  eggs_dir      = eggs
3  logs_dir      = logs
4  items_dir     =
5  jobs_to_keep  = 5
6  dbs_dir       = dbs
7  max_proc      = 0
8  max_proc_per_cpu = 4
9  finished_to_keep = 100
10 poll_interval = 5.0
11 bind_address  = 127.0.0.1
12 http_port     = 6800
13 debug         = off
14 runner        = scrapyd.runner
15 application   = scrapyd.app.application
16 launcher      = scrapyd.launcher.Launcher
17 webroot       = scrapyd.website.Root
18
19 [services]
20 schedule.json  = scrapyd.webservice.Schedule
21 cancel.json    = scrapyd.webservice.Cancel
22 addversion.json = scrapyd.webservice.AddVersion
23 listprojects.json = scrapyd.webservice.ListProjects
24 listversions.json = scrapyd.webservice.ListVersions
25 listspiders.json = scrapyd.webservice.ListSpiders
26 delproject.json = scrapyd.webservice.DeleteProject
27 delversion.json = scrapyd.webservice.DeleteVersion
28 listjobs.json  = scrapyd.webservice.ListJobs
29 daemonstatus.json = scrapyd.webservice.DaemonStatus
30
```

它分为 **Scrapyd** 和 **Services** 两种级别，其中每个参数的作用，在 [Scrapyd 官方文档](#)中有详细的介绍。

## Scrapyd 级配置

- 其中 `.._dir` 均为对应的目录设置，如 `eggs_dir`、`logs_dir`、`items_dir`、`dbs_dir`。
- `jobs_to_keep` - 保存每个爬虫的日志记录数，默认为 5，你可以设置为 500。
- `max_proc` - Scrapyd 最大并发数，默认 0 时使用 `max_proc_per_cpu` 设置的数量。
- `max_proc_per_cpu` - 每个 CPU 最大并发数设置，默认为 4。
- `finished_to_keep` - 在 launcher 中保持的进程数，默认 100。
- `poll_interval` - 轮询队列的间隔，默认 5 秒，可以设置 0.2 或其它 int/float 数值。
- `bind_address` - 默认值 127.0.0.1，只允许本机访问，如果想要开启远程访问改为 0.0.0.0 即可。
- `http_port` - http 端口，默认 6800。
- `debug` - Debug 调试开关。
- `runner`、`application`、`launcher`、`webroot` - 关联的是 Scrapyd 项目代码中编写的对应对象。

以上配置只能写在 Scrapyd 级配置中，如果写错了或者写到其他位置，Scrapyd 在读取配置文件的时候是读取不到的。

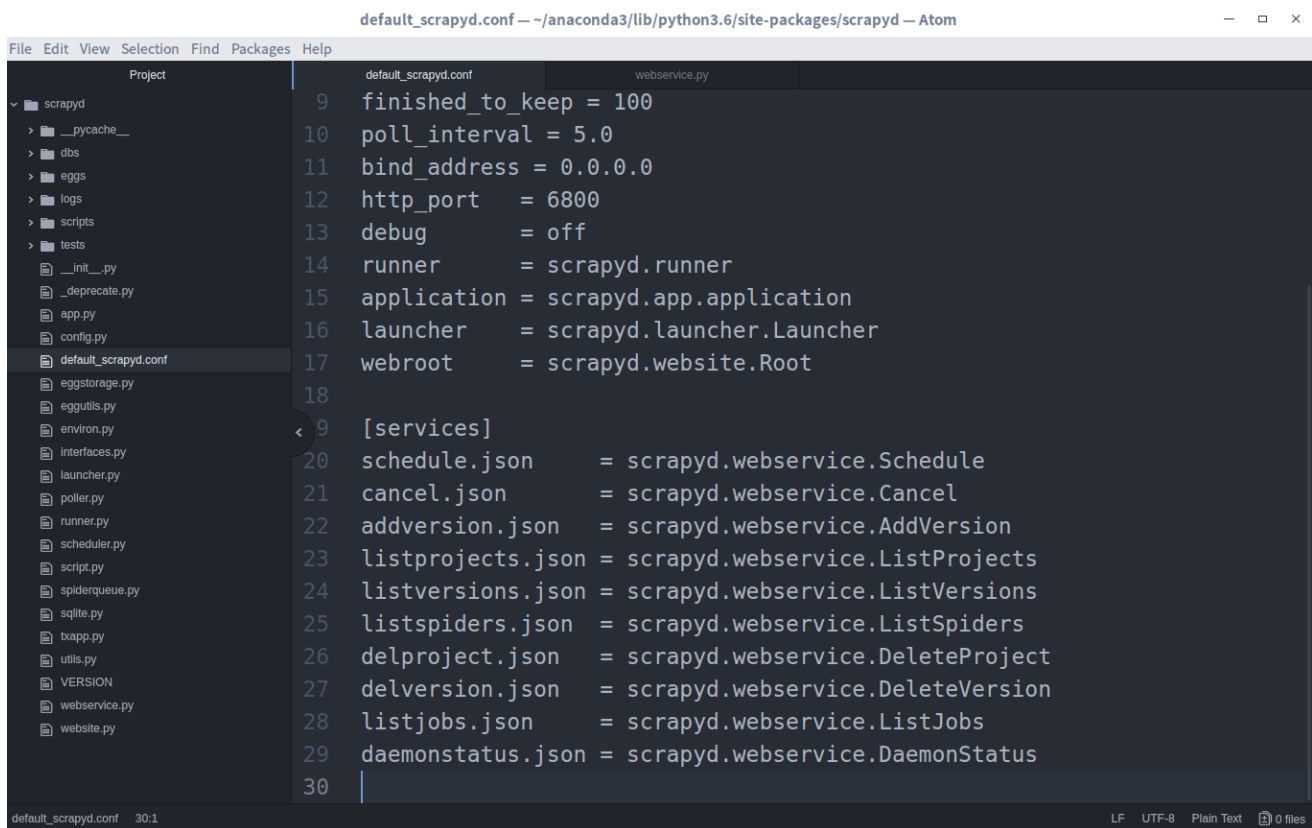
## Services 级配置

Services 级专用于配置 API 的路由映射关系，将 URL 地址与资源关联起来。

假如在 `webservice.py` 中新增一个名为 `Caps` 的类，那么在配置文件中的写法应与其他类的路由写法一致：

```
1 [services]
2 schedule.json      = scrapyd.webservice.Schedule
3 ... ..
4 daemonstatus.json = scrapyd.webservice.DaemonStatus
5 caps.json          = scrapyd.webservice.Caps
6
```

下面这张 GIF 动图演示了整个 API 代码编写、路由配置以及 API 使用的过程。



## 小结

本小节我们学习了 Scrapy 配置文件中各个参数的作用以及配置文件级别的区别，并且通过编写 API 的案例对 Scrapy 配置文件进行了设置。有了实践经验后，我们就可以在后面的小节中对 Scrapy 的配置文件进行自定义设置了。