**@9秒小说阅读服务端开发文档**

**(v1.0.2)**

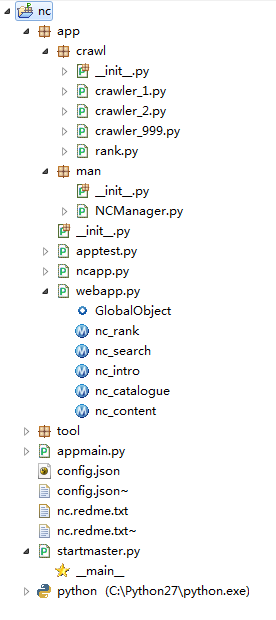
**项目**

|  |  |
| --- | --- |
| **项目名称** | 9秒小说阅读器 分布式服务端 |
| **项目编号** | NC14052800S |
| **当前版本** | v1.0.2 |
| **运行环境** | windows,linux |
| **开发语言** | Python |
| **开发框架** | gfirefly v0.1.5 |
| **作者** | 小新@www.9miao.com |

**更新**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **版本号** | **修订时间** | **修订人** | **操作文件** | **备注** |
| v1.0.1 | 2014.06.03 | 小新 | crawler\_1.py | 修改linux下编码问题 |
| v1.0.2 | 2014.06.23 | 小新 | 本文档 | 添加JSON键说明 |
| v1.0.2 | 2014.06.23 | 小新 | crawler\_1.py | 添加搜索返回100条限制，修改开头空格为全角 |
| V1.0.2 | 2014.06.23 | 大鸡蛋 | 本文档 | 添加了服务器端功能和内容说明 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

1. **项目结构：**



由于9秒小说阅读器的一切数据，均直接从“沸腾文学”上抓取的，所以并不需要开发者自己再上传小说数据，即可直接商业运营。

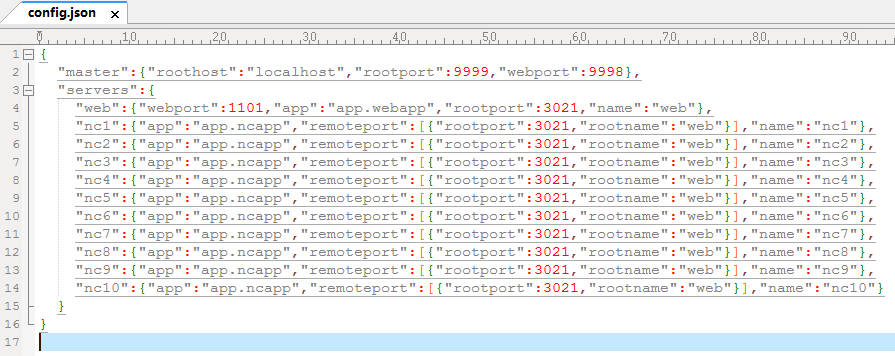
开发者可根据现有服务器爬虫进行修改，再针对其他小说网站制作爬虫，已丰富小说阅读器的内容，为读者提供更好的服务。（欢迎把新制作的爬虫分享到store.9miao.com中，为更多人服务）

此版本9秒小说阅读服务器主要干了3个事情：

1、接收Client的请求；

2、根据收到的请求，抓取 “沸腾文学网站”的小说和列表等相关内容；

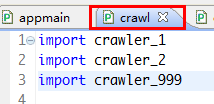
3、对抓取过来的数据，进行解析和重新封包，并返回给Client端；

1. **配置文件**
2. master为整个应用分布式服务器的主进程
3. servers包含了分布式的单个服务，webport为外部访问的端口。其他为内部服务交互接口。
4. 此例中启用了10个节点，可以单独启动，可同时启动，动态检测节点的存活状态。
5. Json键说明：

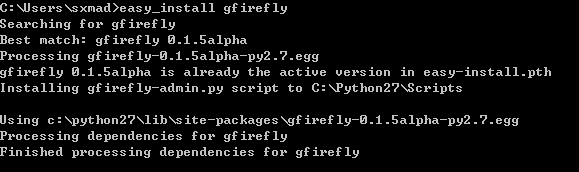
|  |  |
| --- | --- |
| **intro\_url** | 小说信息链接 |
| **cat\_url** | 小说目录链接 |
| **img\_url** | 小说封面链接 |
| **book\_name** | 小说名称 |
| **latest\_chapter** | 最新章节 |
| **book\_author** | 小说作者 |
| **book\_words** | 小说字数 |
| **update\_date** | 更新时间 |
| **book\_source\_site** | 小说源 |
| **book\_chapter** | 章节 |
| **content\_text** | 小说内容与下面的img\_url只取其一 |
| **img\_url** | 小说内容，以图片形式展示，为空则为文字 |

1. **包，模块，类：**
2. app,tool均为firefly默认生成目录。其中app里面包含了应用本身。
3. app/craw此目录包含了所有的爬虫，crawler\_999.py为错误处理，rank为排行，其他crawler\_1.py…….crawler\_n.py(n != 999)均为具体爬虫模块。
4. \_\_init\_\_.py，此项目中，此目录下只做模块导入。
5. man/NCManager.py 此模块做各节点的处理。包括节点状态检测。
6. webapp.py负责接收并转发所有的web请求
7. ncapp.py负责接收webapp转发过来的请求，并调用相应的处理接口。
8. **扩展：**

如果需要扩展添加其他爬虫，只需要在crawl包下添加相应的解析模块即可，如，添加相应的crawler\_2,py, crawler\_3,py, crawler\_4,py…crawler\_n,py并在模块中实现其相应的功能，然后在\_\_init\_\_.py下导入相应模块即可：

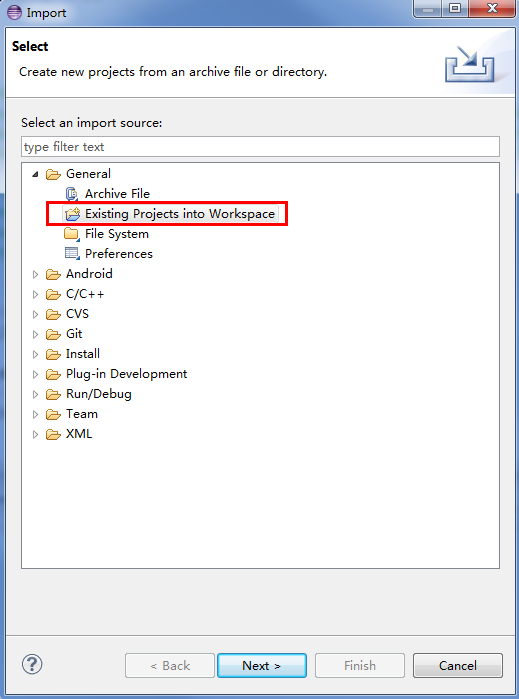


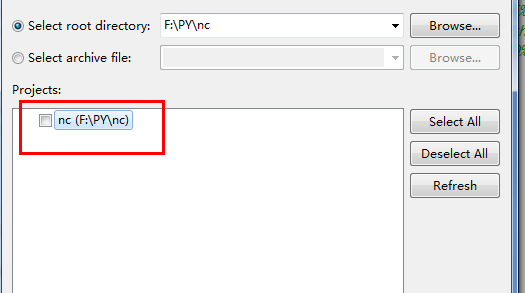
1. **gfirefly环境配置：**
2. 打开cmd，执行命令：easy\_install gfirefly



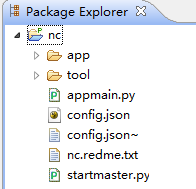
安装过程会把依赖包全部装入，如果期间报错可以独立安装报错项目。

1. eclipse导入工程：

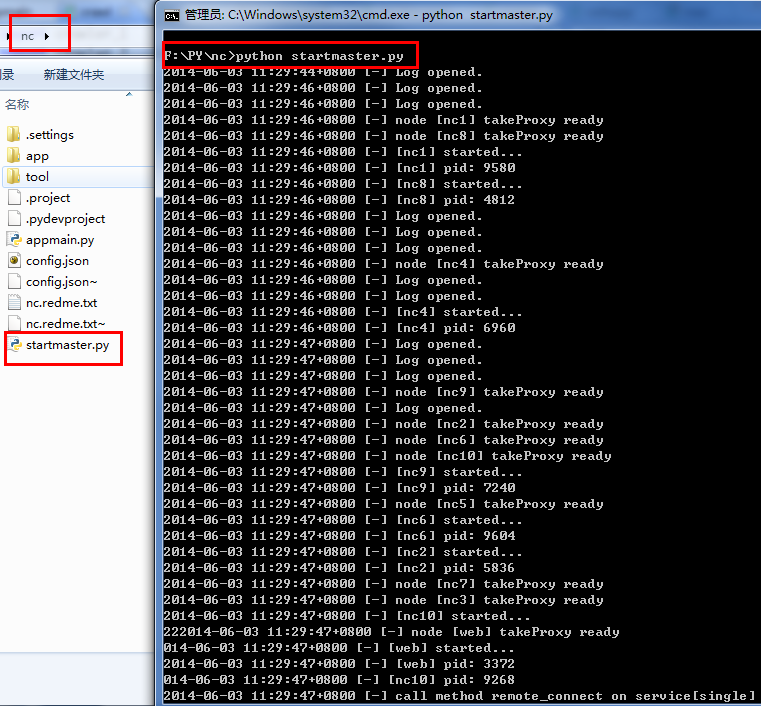




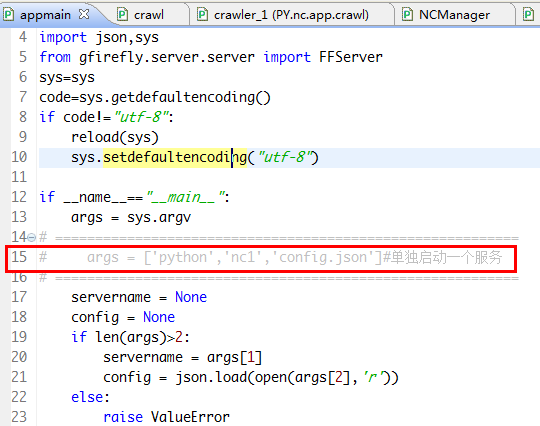
1. 导入成功后：



1. **启动：**
2. 集群启动（不支持调试）：到应用目录下，python startmaster.py



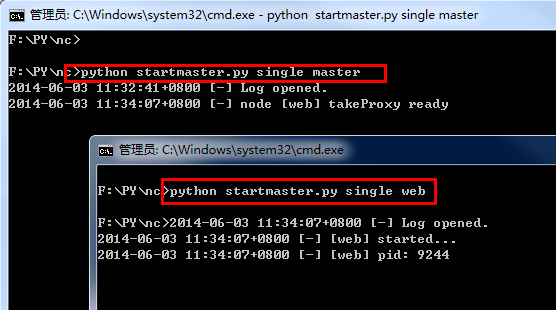
1. 单节点启动（eclipse下支持调试）：
2. 注释掉appmain.py中line 15:



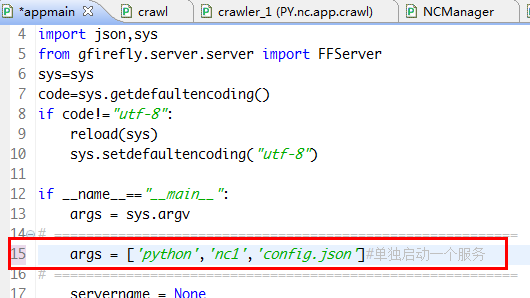
1. nc目录下打开控制台输入：python startmaster.py single master



1. nc目录下，打开新控制台输入：python startmaster.py single web



1. 去掉appmain.py下line15注释：



1. 运行appmain.py

