

READ ME

配置代理IP

Middlewares.py 文件下:

修改proxyUser和proxyPass即可

通行证	proxyUser	通行密钥	proxyPass
	HL4FTH53QV872H8D		41733042BAE0CCE9

```
132 import base64
133
134 # 代理服务器
135 proxyServer = "http://http-dyn.abuyun.com:9020"
136
137 # 代理隧道验证信息
138 proxyUser = "HL4FTH53QV872H8D"
139 proxyPass = "41733042BAE0CCE9"
140
141 # for Python2
142 # proxyAuth = "Basic " + base64.b64encode(proxyUser + ":" + proxyPass)
143 proxyAuth = "Basic " + base64.urlsafe_b64encode(bytes((proxyUser + ":" + proxyPass),
"ascii")).decode("utf8")
```

——配置完成后可运行Spider

运行Spider

```

F:\.
baseInfo.json
fid.json
scrapy.cfg
__init__.py

+---info
|   all.log
|   items.py
|   middlewares.py
|   pipelines.py
|   settings.py
|   start.py
|   user_agents.py
|   __init__.py
|
+---spiders
|   baseSpider.py
|   testSpider.py
|   __init__.py
|
\---__pycache__
|   baseSpider.cpython-37.pyc
|   testSpider.cpython-37.pyc
|   __init__.cpython-37.pyc
|
\---__pycache__
|   items.cpython-37.pyc
|   middlewares.cpython-37.pyc
|   pipelines.cpython-37.pyc
|   settings.cpython-37.pyc
|   user_agents.cpython-37.pyc
|   __init__.cpython-37.pyc
|
\---__pycache__
|   __init__.cpython-37.pyc

```

路径：F:\pycharm\workspace\info\info (* F:\pycharm\workspace\ *为项目所在目录)

指令（终端下）：scrapy crawl base

查看结果

__pycache__	2020/2/15 15:20	文件夹	
info	2020/2/20 10:12	文件夹	
__init__.py	2020/2/15 15:18	JetBrains PyCharm	0 KB
baseInfo.json	2020/2/20 11:19	JSON 文件	538 KB
fid.json	2020/2/16 16:37	JSON 文件	0 KB
scrapy.cfg	2020/2/10 14:58	CFG 文件	1 KB

该目录下baseInfo.json为输出结果（默认输出为不覆盖原文件，可能存在重复结果的现象。）