

```
1 C:\Users\Stepthen\AppData\Local\Programs\
  Python\Python38\python.exe D:\GraduateDesign\
  Code\WeiboSpider-Master\weibospider\
  weibospider\run_spider.py
2 2023-03-10 13:59:39 [scrapy.utils.log] INFO:
  Scrapy 1.5.1 started (bot: spider)
3 2023-03-10 13:59:39 [scrapy.utils.log] INFO:
  Versions: lxml 4.6.3.0, libxml2 2.9.5,
  cssselect 1.2.0, parsel 1.6.0, w3lib 1.22.0,
  Twisted 21.2.0, Python 3.8.10 (tags/v3.8.10:
  3d8993a, May 3 2021, 11:48:03) [MSC v.1928
  64 bit (AMD64)], pyOpenSSL 20.0.1 (OpenSSL 1.
  1.1k 25 Mar 2021), cryptography 3.4.7,
  Platform Windows-10-10.0.19045-SP0
4 2023-03-10 13:59:39 [scrapy.crawler] INFO:
  Overridden settings: {'BOT_NAME': 'spider', '
  CONCURRENT_REQUESTS': 100, 'DOWNLOAD_DELAY':
  0.5, 'DUPEFILTER_CLASS': '
  scrapy_redis_bloomfilter.dupefilter.
  RFPDupeFilter', 'LOG_LEVEL': 'INFO', '
  NEWSPIDER_MODULE': 'spiders', 'SCHEDULER': '
  scrapy_redis_bloomfilter.scheduler.Scheduler
  ', 'SPIDER_MODULES': ['spiders']}
5 2023-03-10 13:59:39 [scrapy.middleware] INFO
  : Enabled extensions:
6 ['scrapy.extensions.corestats.CoreStats', '
  scrapy.extensions.logstats.LogStats']
7 2023-03-10 13:59:39 [search_spider] INFO:
  Reading start URLs from redis key '
  search_spider:start_urls' (batch size: 100,
  encoding: utf-8
8 2023-03-10 13:59:39 [scrapy.middleware] INFO
  : Enabled downloader middlewares:
9 ['middlewares.IPProxyMiddleware',
10 'scrapy.downloadermiddlewares.httpproxy.
```

```
10 HttpProxyMiddleware',
11 'middlewares.RedirectMiddleware',
12 'scrapy.downloadermiddlewares.httppauth.
   HttpAuthMiddleware',
13 'middlewares.CookieMiddleware',
14 'scrapy.downloadermiddlewares.
   downloadtimeout.DownloadTimeoutMiddleware',
15 'scrapy.downloadermiddlewares.defaultheaders.
   DefaultHeadersMiddleware',
16 'scrapy.downloadermiddlewares.useragent.
   UserAgentMiddleware',
17 'scrapy.downloadermiddlewares.retry.
   RetryMiddleware',
18 'scrapy.downloadermiddlewares.redirect.
   MetaRefreshMiddleware',
19 'scrapy.downloadermiddlewares.
   httpcompression.HttpCompressionMiddleware',
20 'scrapy.downloadermiddlewares.stats.
   DownloaderStats']
21 2023-03-10 13:59:39 [scrapy.middleware] INFO
   : Enabled spider middlewares:
22 ['scrapy.spidermiddlewares.httperror.
   HttpErrorMiddleware',
23 'scrapy.spidermiddlewares.offsite.
   OffsiteMiddleware',
24 'scrapy.spidermiddlewares.referer.
   RefererMiddleware',
25 'scrapy.spidermiddlewares.urllength.
   UrlLengthMiddleware',
26 'scrapy.spidermiddlewares.depth.
   DepthMiddleware']
27 2023-03-10 13:59:39 [scrapy.middleware] INFO
   : Enabled item pipelines:
28 ['pipelines.WriterPipeline']
29 2023-03-10 13:59:39 [scrapy.core.engine] INFO
```

```
29 : Spider opened
30 2023-03-10 13:59:39 [scrapy.extensions.
    logstats] INFO: Crawled 0 pages (at 0 pages/
    min), scraped 0 items (at 0 items/min)
31 search_spider:bloomfilter 30 6
32 2023-03-10 14:00:39 [scrapy.extensions.
    logstats] INFO: Crawled 99 pages (at 99 pages
    /min), scraped 1174 items (at 1174 items/min)
33 2023-03-10 14:01:39 [scrapy.extensions.
    logstats] INFO: Crawled 198 pages (at 99
    pages/min), scraped 2340 items (at 1166 items
    /min)
34 2023-03-10 14:02:39 [scrapy.extensions.
    logstats] INFO: Crawled 294 pages (at 96
    pages/min), scraped 3391 items (at 1051 items
    /min)
35 2023-03-10 14:03:39 [scrapy.extensions.
    logstats] INFO: Crawled 390 pages (at 96
    pages/min), scraped 4366 items (at 975 items/
    min)
36 2023-03-10 14:04:39 [scrapy.extensions.
    logstats] INFO: Crawled 488 pages (at 98
    pages/min), scraped 5585 items (at 1219 items
    /min)
37 2023-03-10 14:05:39 [scrapy.extensions.
    logstats] INFO: Crawled 591 pages (at 103
    pages/min), scraped 6766 items (at 1181 items
    /min)
38 2023-03-10 14:06:39 [scrapy.extensions.
    logstats] INFO: Crawled 687 pages (at 96
    pages/min), scraped 7928 items (at 1162 items
    /min)
39 2023-03-10 14:07:39 [scrapy.extensions.
    logstats] INFO: Crawled 736 pages (at 49
    pages/min), scraped 8595 items (at 667 items/
```

```
39 min)
40 2023-03-10 14:09:19 [scrapy.extensions.
    logstats] INFO: Crawled 799 pages (at 63
    pages/min), scraped 9367 items (at 772 items/
    min)
41 2023-03-10 14:09:39 [scrapy.extensions.
    logstats] INFO: Crawled 832 pages (at 33
    pages/min), scraped 9778 items (at 411 items/
    min)
42 2023-03-10 14:10:39 [scrapy.extensions.
    logstats] INFO: Crawled 932 pages (at 100
    pages/min), scraped 11001 items (at 1223
    items/min)
43 2023-03-10 14:11:39 [scrapy.extensions.
    logstats] INFO: Crawled 1029 pages (at 97
    pages/min), scraped 12276 items (at 1275
    items/min)
44 2023-03-10 14:12:39 [scrapy.extensions.
    logstats] INFO: Crawled 1125 pages (at 96
    pages/min), scraped 13455 items (at 1179
    items/min)
45 2023-03-10 14:13:39 [scrapy.extensions.
    logstats] INFO: Crawled 1224 pages (at 99
    pages/min), scraped 14527 items (at 1072
    items/min)
46 2023-03-10 14:14:39 [scrapy.extensions.
    logstats] INFO: Crawled 1233 pages (at 9
    pages/min), scraped 14608 items (at 81 items/
    min)
47 2023-03-10 14:15:39 [scrapy.extensions.
    logstats] INFO: Crawled 1233 pages (at 0
    pages/min), scraped 14608 items (at 0 items/
    min)
48 2023-03-10 14:16:39 [scrapy.extensions.
    logstats] INFO: Crawled 1233 pages (at 0
```

```
48 pages/min), scraped 14608 items (at 0 items/
min)
49 2023-03-10 14:17:37 [scrapy.crawler] INFO:
Received SIGINT, shutting down gracefully.
Send again to force
50 2023-03-10 14:17:37 [scrapy.core.engine] INFO
: Closing spider (shutdown)
51 2023-03-10 14:17:37 [scrapy.statscollectors]
INFO: Dumping Scrapy stats:
52 {'downloader/request_bytes': 932850,
53  'downloader/request_count': 1234,
54  'downloader/request_method_count/GET': 1234,
55  'downloader/response_bytes': 29444778,
56  'downloader/response_count': 1234,
57  'downloader/response_status_count/200': 1233
58  ,
59  'downloader/response_status_count/302': 1,
60  'finish_reason': 'shutdown',
61  'finish_time': datetime.datetime(2023, 3, 10
62  , 6, 17, 37, 740016),
63  'item_scraped_count': 14608,
64  'log_count/INFO': 26,
65  'response_received_count': 1233,
66  'scheduler/dequeued/redis': 1234,
67  'scheduler/enqueued/redis': 1236,
68  'start_time': datetime.datetime(2023, 3, 10
69  , 5, 59, 39, 967317)}
70 2023-03-10 14:17:37 [scrapy.core.engine] INFO
: Spider closed (shutdown)
68
69 Process finished with exit code 0
70
```