(爬虫课件) E:\pycharm\爬虫课件\Scripts\Eight-scrapy\middlePro\middlePro>scrapy crawl middle

2020-09-16 16:47:53 [scrapy.utils.log] INFO: Scrapy 2.3.0 started (bot: middlePro)

2020-09-16 16:47:53 [scrapy.utils.log] INFO: Versions: lxml 4.5.2.0, libxml2 2.9.5, cssselect

1.1.0, parsel 1.6.0, w3lib 1.22.0, Twisted 20.3.0, Python 3.7.9 (tags/v3.7.9:13c94747c7, Aug

17 2020, 18:58:18) [MSC v.1900 64 bit (AMD64)], pyOpenSSL 19.1.0 (OpenSSL 1.1.1g 21 Apr 202

0), cryptography 3.1, Platform Windows-10-10.0.18362-SP0

2020-09-16 16:47:53 [scrapy.utils.log] DEBUG: Using reactor: twisted.internet.selectreactor.S

electReactor

2020-09-16 16:47:54 [scrapy.crawler] INFO: Overridden settings:

{'BOT\_NAME': 'middlePro',

'NEWSPIDER\_MODULE': 'middlePro.spiders',

'SPIDER\_MODULES': ['middlePro.spiders']}

2020-09-16 16:47:54 [scrapy.extensions.telnet] INFO: Telnet Password: ed3be3e02185a8d5

2020-09-16 16:47:54 [scrapy.middleware] INFO: Enabled extensions:

['scrapy.extensions.corestats.CoreStats',

'scrapy.extensions.telnet.TelnetConsole',

'scrapy.extensions.logstats.LogStats']

2020-09-16 16:47:54 [scrapy.middleware] INFO: Enabled downloader middlewares:

['scrapy.downloadermiddlewares.httpauth.HttpAuthMiddleware',

'scrapy.downloadermiddlewares.downloadtimeout.DownloadTimeoutMiddleware',

'scrapy.downloadermiddlewares.defaultheaders.DefaultHeadersMiddleware',

'scrapy.downloadermiddlewares.useragent.UserAgentMiddleware',

'middlePro.middlewares.MiddleproDownloaderMiddleware',

'scrapy.downloadermiddlewares.retry.RetryMiddleware',

'scrapy.downloadermiddlewares.redirect.MetaRefreshMiddleware',

'scrapy.downloadermiddlewares.httpcompression.HttpCompressionMiddleware',

'scrapy.downloadermiddlewares.redirect.RedirectMiddleware',

'scrapy.downloadermiddlewares.cookies.CookiesMiddleware',

'scrapy.downloadermiddlewares.httpproxy.HttpProxyMiddleware',

'scrapy.downloadermiddlewares.stats.DownloaderStats']

2020-09-16 16:47:54 [scrapy.middleware] INFO: Enabled spider middlewares:

['scrapy.spidermiddlewares.httperror.HttpErrorMiddleware',

'scrapy.spidermiddlewares.offsite.OffsiteMiddleware',

'scrapy.spidermiddlewares.referer.RefererMiddleware',

'scrapy.spidermiddlewares.urllength.UrlLengthMiddleware',

'scrapy.spidermiddlewares.depth.DepthMiddleware']

2020-09-16 16:47:54 [scrapy.middleware] INFO: Enabled item pipelines:

[]

2020-09-16 16:47:54 [scrapy.core.engine] INFO: Spider opened

2020-09-16 16:47:54 [scrapy.extensions.logstats] INFO: Crawled 0 pages (at 0 pages/min), scra

ped 0 items (at 0 items/min)

2020-09-16 16:47:54 [scrapy.extensions.telnet] INFO: Telnet console listening on 127.0.0.1:60

23

2020-09-16 16:48:15 [scrapy.downloadermiddlewares.retry] DEBUG: Retrying <GET http://www.baid

u.com/s?wd=ip> (failed 1 times): TCP connection timed out: 10060: 由于连接方在一段时间后没有

正确答复或连接的主机没有反应，连接尝试失败。.

2020-09-16 16:48:36 [scrapy.downloadermiddlewares.retry] DEBUG: Retrying <GET http://www.baid

u.com/s?wd=ip> (failed 2 times): TCP connection timed out: 10060: 由于连接方在一段时间后没有

正确答复或连接的主机没有反应，连接尝试失败。.

2020-09-16 16:53:54 [scrapy.extensions.logstats] INFO: Crawled 0 pages (at 0 pages/min), scra

ped 0 items (at 0 items/min)

2020-09-16 16:53:55 [scrapy.crawler] INFO: Received SIGINT, shutting down gracefully. Send ag

ain to force

2020-09-16 16:53:55 [scrapy.core.engine] INFO: Closing spider (shutdown)

2020-09-16 16:53:55 [scrapy.crawler] INFO: Received SIGINT twice, forcing unclean shutdown

2020-09-16 16:53:55 [scrapy.downloadermiddlewares.retry] ERROR: Gave up retrying <GET http://

www.baidu.com/s?wd=ip> (failed 3 times): An error occurred while connecting: [Failure instanc

e: Traceback (failure with no frames): <class 'twisted.internet.error.ConnectionLost'>: Conne

ction to the other side was lost in a non-clean fashion: Connection lost.

].

2020-09-16 16:53:55 [scrapy.statscollectors] INFO: Dumping Scrapy stats:

{'downloader/exception\_count': 18,

'downloader/exception\_type\_count/twisted.internet.error.ConnectError': 1,

'downloader/exception\_type\_count/twisted.internet.error.TCPTimedOutError': 17,

'downloader/request\_bytes': 5202,

'downloader/request\_count': 18,

'downloader/request\_method\_count/GET': 18,

'elapsed\_time\_seconds': 361.092988,

'finish\_reason': 'shutdown',

'finish\_time': datetime.datetime(2020, 9, 16, 8, 53, 55, 776789),

'log\_count/DEBUG': 2,

'log\_count/ERROR': 16,

'log\_count/INFO': 18,

'retry/count': 2,

'retry/max\_reached': 16,

'retry/reason\_count/twisted.internet.error.TCPTimedOutError': 2,

'scheduler/dequeued': 18,

'scheduler/dequeued/memory': 18,

'scheduler/enqueued': 19,

'scheduler/enqueued/memory': 19,

'start\_time': datetime.datetime(2020, 9, 16, 8, 47, 54, 683801)}

2020-09-16 16:53:55 [scrapy.core.engine] INFO: Spider closed (shutdown)