

[文件](#) » 标杆

基准测试

版本0.17中的新功能。

Scrapy附带一个简单的基准测试套件，可以生成本地HTTP服务器并以尽可能高的速度对其进行爬网。此基准测试的目标是了解Scrapy如何在您的硬件中执行，以便有一个共同的比较基线。它使用一个简单的蜘蛛，什么都不做，只是跟随链接。

要运行它，请使用：

```
scrapy bench
```

你应该看到这样的输出：

```
2016-12-16 21:18:48 [scrapy.utils.log] INFO: Scrapy 1.2.2 started (bot: quotesbot)
2016-12-16 21:18:48 [scrapy.utils.log] INFO: Overridden settings: {'CLOSESPIDER_TIMEOUT': 10,
'ROBOTSTXT_OBEY': True, 'SPIDER_MODULES': ['quotesbot.spiders'], 'LOGSTATS_INTERVAL': 1,
'BOT_NAME': 'quotesbot', 'LOG_LEVEL': 'INFO', 'NEWSPIDER_MODULE': 'quotesbot.spiders'}
2016-12-16 21:18:49 [scrapy.middleware] INFO: Enabled extensions:
['scrapy.extensions.closespider.CloseSpider',
'scrapy.extensions.logstats.LogStats',
'scrapy.extensions.telnet.TelnetConsole',
'scrapy.extensions.corestats.CoreStats']
2016-12-16 21:18:49 [scrapy.middleware] INFO: Enabled downloader middlewares:
['scrapy.downloadermiddlewares.robotstxt.RobotsTxtMiddleware',
'scrapy.downloadermiddlewares.httpauth.HttpAuthMiddleware',
'scrapy.downloadermiddlewares.downloadtimeout.DownloadTimeoutMiddleware',
'scrapy.downloadermiddlewares.defaultheaders.DefaultHeadersMiddleware',
'scrapy.downloadermiddlewares.useragent.UserAgentMiddleware',
'scrapy.downloadermiddlewares.retry.RetryMiddleware',
'scrapy.downloadermiddlewares.redirect.MetaRefreshMiddleware',
'scrapy.downloadermiddlewares.httpcompression.HttpCompressionMiddleware',
'scrapy.downloadermiddlewares.redirect.RedirectMiddleware',
'scrapy.downloadermiddlewares.cookies.CookiesMiddleware',
'scrapy.downloadermiddlewares.stats.DownloaderStats']
2016-12-16 21:18:49 [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']
2016-12-16 21:18:49 [scrapy.middleware] INFO: Enabled item pipelines:
[]
2016-12-16 21:18:49 [scrapy.core.engine] INFO: Spider opened
2016-12-16 21:18:49 [scrapy.extensions.logstats] INFO: Crawled 0 pages (at 0 pages/min),
scraped 0 items (at 0 items/min)
2016-12-16 21:18:50 [scrapy.extensions.logstats] INFO: Crawled 70 pages (at 4200 pages/min),
scraped 0 items (at 0 items/min)
2016-12-16 21:18:51 [scrapy.extensions.logstats] INFO: Crawled 134 pages (at 3840 pages/min),
scraped 0 items (at 0 items/min)
2016-12-16 21:18:52 [scrapy.extensions.logstats] INFO: Crawled 198 pages (at 3840 pages/min),
scraped 0 items (at 0 items/min)
2016-12-16 21:18:53 [scrapy.extensions.logstats] INFO: Crawled 254 pages (at 3360 pages/min),
scraped 0 items (at 0 items/min)
2016-12-16 21:18:54 [scrapy.extensions.logstats] INFO: Crawled 302 pages (at 2880 pages/min),
scraped 0 items (at 0 items/min)
2016-12-16 21:18:55 [scrapy.extensions.logstats] INFO: Crawled 358 pages (at 3360 pages/min),
scraped 0 items (at 0 items/min)
2016-12-16 21:18:56 [scrapy.extensions.logstats] INFO: Crawled 406 pages (at 2880 pages/min),
scraped 0 items (at 0 items/min)
2016-12-16 21:18:57 [scrapy.extensions.logstats] INFO: Crawled 438 pages (at 1920 pages/min),
scraped 0 items (at 0 items/min)
2016-12-16 21:18:58 [scrapy.extensions.logstats] INFO: Crawled 470 pages (at 1920 pages/min),
scraped 0 items (at 0 items/min)
2016-12-16 21:18:59 [scrapy.core.engine] INFO: Closing spider (closespider_timeout)
2016-12-16 21:18:59 [scrapy.extensions.logstats] INFO: Crawled 518 pages (at 2880 pages/min),
scraped 0 items (at 0 items/min)
2016-12-16 21:19:00 [scrapy.statscollectors] INFO: Dumping Scrapy stats:
{'downloader/request_bytes': 229995,
'downloader/request_count': 534,
'downloader/request_method_count/GET': 534,
'downloader/response_bytes': 1565504,
'downloader/response_count': 534,
'downloader/response_status_count/200': 534,
'finish_reason': 'closespider_timeout',
'finish_time': datetime.datetime(2016, 12, 16, 19, 0, 647725),
'log_count/INFO': 17,
'request_depth_max': 19,
'response_received_count': 534,
'scheduler/dequeued': 533,
'scheduler/dequeued/memory': 533,
'scheduler/enqueued': 10661,
'scheduler/enqueued/memory': 10661,
'start_time': datetime.datetime(2016, 12, 16, 18, 49, 799869)}
2016-12-16 21:19:00 [scrapy.core.engine] INFO: Spider closed (closespider_timeout)
```

这告诉您Scrapy能够在您运行它的硬件中每分钟爬行大约3000页。请注意，这是一个非常简单的蜘蛛，旨在跟踪链接，您编写的任何自定义蜘蛛可能会做更多的东西，导致较慢的爬行率。速度有多慢取决于你的蜘蛛的作用程度以及它的编写效果。

将来，更多案例将添加到基准测试套件中，以涵盖其他常见场景。