

AutoThrottle extension

This is an extension for automatically throttling crawling speed based on load of both the Scrapy server and the website you are crawling.

Design goals

1. be nicer to sites instead of using default download delay of zero
2. automatically adjust Scrapy to the optimum crawling speed, so the user doesn't have to tune the download delays to find the optimum one. The user only needs to specify the maximum concurrent requests it allows, and the extension does the rest.

How it works

AutoThrottle extension adjusts download delays dynamically to make spider send `:setting:'AUTOTHROTTLE_TARGET_CONCURRENCY'` concurrent requests on average to each remote website.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]autothrottle.rst, line 24); [backlink](#)

Unknown interpreted text role "setting".

It uses download latency to compute the delays. The main idea is the following: if a server needs `latency` seconds to respond, a client should send a request each `latency/N` seconds to have `N` requests processed in parallel.

Instead of adjusting the delays one can just set a small fixed download delay and impose hard limits on concurrency using `:setting:'CONCURRENT_REQUESTS_PER_DOMAIN'` or `:setting:'CONCURRENT_REQUESTS_PER_IP'` options. It will provide a similar effect, but there are some important differences:

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]autothrottle.rst, line 33); [backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]autothrottle.rst, line 33); [backlink](#)

Unknown interpreted text role "setting".

- because the download delay is small there will be occasional bursts of requests;
- often non-200 (error) responses can be returned faster than regular responses, so with a small download delay and a hard concurrency limit crawler will be sending requests to server faster when server starts to return errors. But this is an opposite of what crawler should do - in case of errors it makes more sense to slow down: these errors may be caused by the high request rate.

AutoThrottle doesn't have these issues.

Throttling algorithm

AutoThrottle algorithm adjusts download delays based on the following rules:

1. spiders always start with a download delay of `:setting:'AUTOTHROTTLE_START_DELAY';`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]autothrottle.rst, line 55); [backlink](#)

Unknown interpreted text role "setting".

2. when a response is received, the target download delay is calculated as `latency / N` where `latency` is a latency of the response, and `N` is `:setting:'AUTOTHROTTLE_TARGET_CONCURRENCY'`.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]autothrottle.rst, line 57); [backlink](#)

Unknown interpreted text role "setting".

- download delay for next requests is set to the average of previous download delay and the target download delay;
- latencies of non-200 responses are not allowed to decrease the delay;
- download delay can't become less than `:setting:'DOWNLOAD_DELAY'` or greater than `:setting:'AUTOTHROTTLE_MAX_DELAY'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]autothrottle.rst, line 63); [backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]autothrottle.rst, line 63); [backlink](#)

Unknown interpreted text role "setting".

Note

The AutoThrottle extension honours the standard Scrapy settings for concurrency and delay. This means that it will respect `:setting:'CONCURRENT_REQUESTS_PER_DOMAIN'` and `:setting:'CONCURRENT_REQUESTS_PER_IP'` options and never set a download delay lower than `:setting:'DOWNLOAD_DELAY'`.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]autothrottle.rst, line 66); [backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]autothrottle.rst, line 66); [backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]autothrottle.rst, line 66); [backlink](#)

Unknown interpreted text role "setting".

In Scrapy, the download latency is measured as the time elapsed between establishing the TCP connection and receiving the HTTP headers.

Note that these latencies are very hard to measure accurately in a cooperative multitasking environment because Scrapy may be busy processing a spider callback, for example, and unable to attend downloads. However, these latencies should still give a reasonable estimate of how busy Scrapy (and ultimately, the server) is, and this extension builds on that premise.

Settings

The settings used to control the AutoThrottle extension are:

- `:setting:'AUTOTHROTTLE_ENABLED'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]autothrottle.rst, line 88); [backlink](#)

Unknown interpreted text role "setting".

- `:setting:'AUTOTHROTTLE_START_DELAY'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics] autothrottle.rst, line 89); [backlink](#)

Unknown interpreted text role "setting".

- :setting:`AUTOTHROTTLE_MAX_DELAY`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics] autothrottle.rst, line 90); [backlink](#)

Unknown interpreted text role "setting".

- :setting:`AUTOTHROTTLE_TARGET_CONCURRENCY`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics] autothrottle.rst, line 91); [backlink](#)

Unknown interpreted text role "setting".

- :setting:`AUTOTHROTTLE_DEBUG`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics] autothrottle.rst, line 92); [backlink](#)

Unknown interpreted text role "setting".

- :setting:`CONCURRENT_REQUESTS_PER_DOMAIN`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics] autothrottle.rst, line 93); [backlink](#)

Unknown interpreted text role "setting".

- :setting:`CONCURRENT_REQUESTS_PER_IP`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics] autothrottle.rst, line 94); [backlink](#)

Unknown interpreted text role "setting".

- :setting:`DOWNLOAD_DELAY`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics] autothrottle.rst, line 95); [backlink](#)

Unknown interpreted text role "setting".

For more information see :ref:`autothrottle-algorithm`.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics] autothrottle.rst, line 97); [backlink](#)

Unknown interpreted text role "ref".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics] autothrottle.rst, line 99)

Unknown directive type "setting".

```
.. setting:: AUTOTHROTTLE_ENABLED
```

AUTOTHROTTLE_ENABLED

Default: False

Enables the AutoThrottle extension.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics] autothrottle.rst, line 108)

Unknown directive type "setting".

```
.. setting:: AUTOTHROTTLE_START_DELAY
```

AUTOTHROTTLE_START_DELAY

Default: 5.0

The initial download delay (in seconds).

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics] autothrottle.rst, line 117)

Unknown directive type "setting".

```
.. setting:: AUTOTHROTTLE_MAX_DELAY
```

AUTOTHROTTLE_MAX_DELAY

Default: 60.0

The maximum download delay (in seconds) to be set in case of high latencies.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics] autothrottle.rst, line 126)

Unknown directive type "setting".

```
.. setting:: AUTOTHROTTLE_TARGET_CONCURRENCY
```

AUTOTHROTTLE_TARGET_CONCURRENCY

Default: 1.0

Average number of requests Scrapy should be sending in parallel to remote websites.

By default, AutoThrottle adjusts the delay to send a single concurrent request to each of the remote websites. Set this option to a higher value (e.g. 2.0) to increase the throughput and the load on remote servers. A lower `AUTOTHROTTLE_TARGET_CONCURRENCY` value (e.g. 0.5) makes the crawler more conservative and polite.

Note that `:setting: CONCURRENT_REQUESTS_PER_DOMAIN` and `:setting: CONCURRENT_REQUESTS_PER_IP` options are still respected when AutoThrottle extension is enabled. This means that if `AUTOTHROTTLE_TARGET_CONCURRENCY` is set to a value higher than `:setting: CONCURRENT_REQUESTS_PER_DOMAIN` or `:setting: CONCURRENT_REQUESTS_PER_IP`, the crawler won't reach this number of concurrent requests.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics] autothrottle.rst, line 142); [backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics] autothrottle.rst, line 142); [backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics] autothrottle.rst, line 142); [backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics] autothrottle.rst, line 142); [backlink](#)

Unknown interpreted text role "setting".

At every given time point Scrapy can be sending more or less concurrent requests than `AUTOTHROTTLE_TARGET_CONCURRENCY`; it is a suggested value the crawler tries to approach, not a hard limit.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics] autothrottle.rst, line 154)

Unknown directive type "setting".

```
.. setting:: AUTOTHROTTLE_DEBUG
```

AUTOTHROTTLE_DEBUG

Default: `False`

Enable `AutoThrottle` debug mode which will display stats on every response received, so you can see how the throttling parameters are being adjusted in real time.