

Downloader Middleware

The downloader middleware is a framework of hooks into Scrapy's request/response processing. It's a light, low-level system for globally altering Scrapy's requests and responses.

Activating a downloader middleware

To activate a downloader middleware component, add it to the `setting:'DOWNLOADER_MIDDLEWARES'` setting, which is a dict whose keys are the middleware class paths and their values are the middleware orders.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 16);
[backlink](#)

Unknown interpreted text role "setting".

Here's an example:

```
DOWNLOADER_MIDDLEWARES = {
    'myproject.middlewares.CustomDownloaderMiddleware': 543,
}
```

The `setting:'DOWNLOADER_MIDDLEWARES'` setting is merged with the `setting:'DOWNLOADER_MIDDLEWARES_BASE'` setting defined in Scrapy (and not meant to be overridden) and then sorted by order to get the final sorted list of enabled middlewares: the first middleware is the one closer to the engine and the last is the one closer to the downloader. In other words, the `meth:~scrapy.downloadermiddlewares.DownloaderMiddleware.process_request` method of each middleware will be invoked in increasing middleware order (100, 200, 300, ...) and the `meth:~scrapy.downloadermiddlewares.DownloaderMiddleware.process_response` method of each middleware will be invoked in decreasing order.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 26);
[backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 26);
[backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 26);
[backlink](#)

Unknown interpreted text role "meth".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 26);
[backlink](#)

Unknown interpreted text role "meth".

To decide which order to assign to your middleware see the `setting:'DOWNLOADER_MIDDLEWARES_BASE'` setting and pick a value according to where you want to insert the middleware. The order does matter because each middleware performs a different action and your middleware could depend on some previous (or subsequent) middleware being applied.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 36);
[backlink](#)

Unknown interpreted text role "setting".

If you want to disable a built-in middleware (the ones defined in `setting:'DOWNLOADER_MIDDLEWARES_BASE'` and enabled by default) you must define it in your project's `setting:'DOWNLOADER_MIDDLEWARES'` setting and assign `None` as its value. For example, if you want to disable the user-agent middleware:

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 42);
[backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 42);
[backlink](#)

Unknown interpreted text role "setting".

```
DOWNLOADER_MIDDLEWARES = {
    'myproject.middlewares.CustomDownloaderMiddleware': 543,
    'scrapy.downloadermiddlewares.useragent.UserAgentMiddleware': None,
}
```

Finally, keep in mind that some middlewares may need to be enabled through a particular setting. See each middleware documentation for more info.

Writing your own downloader middleware

Each downloader middleware is a Python class that defines one or more of the methods defined below.

The main entry point is the `from_crawler` class method, which receives a `:class:`~scrapy.crawler.Crawler`` instance. The `:class:`~scrapy.crawler.Crawler`` object gives you access, for example, to the `ref` settings <topics-settings>`.

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

Unknown interpreted text role "class".

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

Unknown interpreted text role "class".

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

Unknown interpreted text role "ref".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 67)

Unknown directive type "module".

```
.. module:: scrapy.downloadermiddlewares
```

Note

Any of the downloader middleware methods may also return a deferred.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 73)

Unknown directive type "method".

```
.. method:: process_request(request, spider)
```

This method is called for each request that goes through the download middleware.

`:meth:`process_request`` should either: return ``None``, return a `:class:`~scrapy.Response`` object, return a `:class:`~scrapy.http.Request`` object, or raise `:exc:`~scrapy.exceptions.IgnoreRequest``.

If it returns ``None``, Scrapy will continue processing this request, executing all other middlewares until, finally, the appropriate downloader handler is called the request performed (and its response downloaded).

If it returns a `:class:`~scrapy.http.Response`` object, Scrapy won't bother calling *any* other `:meth:`process_request`` or `:meth:`process_exception`` methods, or the appropriate download function; it'll return that response. The `:meth:`process_response`` methods of installed middleware is always called on every response.

If it returns a `:class:`~scrapy.Request`` object, Scrapy will stop calling `:meth:`process_request`` methods and reschedule the returned request. Once the newly returned request is performed, the appropriate middleware chain will be called on the downloaded response.

If it raises an `:exc:`~scrapy.exceptions.IgnoreRequest`` exception, the `:meth:`process_exception`` methods of installed downloader middleware will be called. If none of them handle the exception, the `errback` function of the request (``Request.errback``) is called. If no code handles the raised exception, it is ignored and not logged (unlike other exceptions).

`:param request:` the request being processed
`:type request:` `:class:`~scrapy.Request`` object

`:param spider:` the spider for which this request is intended
`:type spider:` `:class:`~scrapy.Spider`` object

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

Unknown directive type "method".

```
.. method:: process_response(request, response, spider)
```

`:meth:`process_response`` should either: return a `:class:`~scrapy.http.Response`` object, return a `:class:`~scrapy.Request`` object or raise a `:exc:`~scrapy.exceptions.IgnoreRequest`` exception.

If it returns a `:class:`~scrapy.http.Response`` (it could be the same given response, or a brand-new one), that response will continue to be processed with the `:meth:`process_response`` of the next middleware in the chain.

If it returns a `:class:`~scrapy.Request`` object, the middleware chain is halted and the returned request is rescheduled to be downloaded in the future. This is the same behavior as if a request is returned from `:meth:`process_request``.

If it raises an `:exc::~scrapy.exceptions.IgnoreRequest`` exception, the `errback` function of the request (``Request.errback``) is called. If no code handles the raised exception, it is ignored and not logged (unlike other exceptions).

`:param request:` the request that originated the response
`:type request:` is a `:class::~scrapy.Request`` object

`:param response:` the response being processed
`:type response:` `:class::~scrapy.http.Response`` object

`:param spider:` the spider for which this response is intended
`:type spider:` `:class::~scrapy.Spider`` object

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 135)

Unknown directive type "method".

```
.. method:: process_exception(request, exception, spider)

    Scrapy calls :meth:`process_exception` when a download handler
    or a :meth:`process_request` (from a downloader middleware) raises an
    exception (including an :exc:`~scrapy.exceptions.IgnoreRequest` exception)

    :meth:`process_exception` should return: either ``None``,
    a :class:`~scrapy.http.Response` object, or a :class:`~scrapy.Request` object.

    If it returns ``None``, Scrapy will continue processing this exception,
    executing any other :meth:`process_exception` methods of installed middleware,
    until no middleware is left and the default exception handling kicks in.

    If it returns a :class:`~scrapy.http.Response` object, the :meth:`process_response`
    method chain of installed middleware is started, and Scrapy won't bother calling
    any other :meth:`process_exception` methods of middleware.

    If it returns a :class:`~scrapy.Request` object, the returned request is
    rescheduled to be downloaded in the future. This stops the execution of
    :meth:`process_exception` methods of the middleware the same as returning a
    response would.

    :param request: the request that generated the exception
    :type request: is a :class:`~scrapy.Request` object

    :param exception: the raised exception
    :type exception: an ``Exception`` object

    :param spider: the spider for which this request is intended
    :type spider: :class:`~scrapy.Spider` object
```

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 166)

Unknown directive type "method".

```
.. method:: from_crawler(cls, crawler)

    If present, this classmethod is called to create a middleware instance
    from a :class:`~scrapy.crawler.Crawler`. It must return a new instance
    of the middleware. Crawler object provides access to all Scrapy core
    components like settings and signals; it is a way for middleware to
    access them and hook its functionality into Scrapy.

    :param crawler: crawler that uses this middleware
    :type crawler: :class:`~scrapy.crawler.Crawler` object
```

Built-in downloader middleware reference

This page describes all downloader middleware components that come with Scrapy. For information on how to use them and how to write your own downloader middleware, see the [ref: downloader middleware usage guide <topics-downloader-middleware>](#).

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 182);
[backlink](#)

Unknown interpreted text role "ref".

For a list of the components enabled by default (and their orders) see the `:setting:`DOWNLOADER_MIDDLEWARES_BASE`` setting.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 187);
[backlink](#)

Unknown interpreted text role "setting".

CookiesMiddleware

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 195)

Unknown directive type "module".

```
.. module:: scrapy.downloadermiddlewares.cookies
    :synopsis: Cookies Downloader Middleware
```

This middleware enables working with sites that require cookies, such as those that use sessions. It keeps track of cookies sent by web servers, and sends them back on subsequent requests (from that spider), just like web browsers do.

Caution!

When non-UTF8 encoded byte sequences are passed to a `:class:`~scrapy.Request``, the `CookiesMiddleware` will log a warning. Refer to [ref: topics-logging-advanced-customization](#) to customize the logging behaviour.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 205);
[backlink](#)

Unknown interpreted text role "class".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-

resources\scrapy-master\docs\topics\ (scrapy-master) (docs)
(topics) downloader-middleware.rst, line 205); [backlink](#)

Unknown interpreted text role "ref".

Caution!

Cookies set via the `Cookie` header are not considered by the `ref: cookies-mw`. If you need to set cookies for a request, use the `class: Request.cookies <scrapy.Request>` parameter. This is a known current limitation that is being worked on.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\ (scrapy-master) (docs)
(topics) downloader-middleware.rst, line 210); [backlink](#)

Unknown interpreted text role "ref".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\ (scrapy-master) (docs)
(topics) downloader-middleware.rst, line 210); [backlink](#)

Unknown interpreted text role "class".

The following settings can be used to configure the cookie middleware:

- `setting: COOKIES_ENABLED`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\ (scrapy-master) (docs) (topics) downloader-middleware.rst, line 217); [backlink](#)

Unknown interpreted text role "setting".

- `setting: COOKIES_DEBUG`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\ (scrapy-master) (docs) (topics) downloader-middleware.rst, line 218); [backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\ (scrapy-master) (docs) (topics) downloader-middleware.rst, line 220)

Unknown directive type "reqmeta".

```
.. reqmeta:: cookiejar
```

Multiple cookie sessions per spider

There is support for keeping multiple cookie sessions per spider by using the `reqmeta: cookiejar` Request meta key. By default it uses a single cookie jar (session), but you can pass an identifier to use different ones.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\ (scrapy-master) (docs) (topics) downloader-middleware.rst, line 225); [backlink](#)

Unknown interpreted text role "reqmeta".

For example:

```
for i, url in enumerate(urls):
    yield scrapy.Request(url, meta={'cookiejar': i},
        callback=self.parse_page)
```

Keep in mind that the `reqmeta: cookiejar` meta key is not "sticky". You need to keep passing it along on subsequent requests. For example:

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\ (scrapy-master) (docs) (topics) downloader-middleware.rst, line 235); [backlink](#)

Unknown interpreted text role "reqmeta".

```
def parse_page(self, response):
    # do some processing
    return scrapy.Request("http://www.example.com/otherpage",
        meta={'cookiejar': response.meta['cookiejar']},
        callback=self.parse_other_page)
```

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\ (scrapy-master) (docs) (topics) downloader-middleware.rst, line 244)

Unknown directive type "setting".

```
.. setting:: COOKIES_ENABLED
```

COOKIES_ENABLED

Default: True

Whether to enable the cookies middleware. If disabled, no cookies will be sent to web servers.

Notice that despite the value of `setting: COOKIES_ENABLED` setting if `Request.reqmeta: meta[dont_merge_cookies]`

`<dont_merge_cookies>` evaluates to `True` the request cookies will **not** be sent to the web server and received cookies in `:class:'~scrapy.http.Response'` will **not** be merged with the existing cookies.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 254);
[backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 254);
[backlink](#)

Unknown interpreted text role "reqmeta".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 254);
[backlink](#)

Unknown interpreted text role "class".

For more detailed information see the `cookies` parameter in `:class:'~scrapy.Request'`.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 260);
[backlink](#)

Unknown interpreted text role "class".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 263)

Unknown directive type "setting".

```
.. setting:: COOKIES_DEBUG
```

COOKIES_DEBUG

Default: `False`

If enabled, Scrapy will log all cookies sent in requests (i.e. `Cookie` header) and all cookies received in responses (i.e. `Set-Cookie` header).

Here's an example of a log with `:setting:'COOKIES_DEBUG'` enabled:

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 273);
[backlink](#)

Unknown interpreted text role "setting".

```
2011-04-06 14:35:10-0300 [scrapy.core.engine] INFO: Spider opened
2011-04-06 14:35:10-0300 [scrapy.downloadermiddlewares.cookies] DEBUG: Sending cookies to: <GET http://www.diningcity.com/netl
  Cookie: clientlanguage_nl=en_EN
2011-04-06 14:35:14-0300 [scrapy.downloadermiddlewares.cookies] DEBUG: Received cookies from: <200 http://www.diningcity.com/r
  Set-Cookie: JSESSIONID=B-FA4DC0C496C8762AE4F1A620EAB34F38; Path=/
  Set-Cookie: ip_isocode=US
  Set-Cookie: clientlanguage_nl=en_EN; Expires=Thu, 07-Apr-2011 21:21:34 GMT; Path=/
2011-04-06 14:49:50-0300 [scrapy.core.engine] DEBUG: Crawled (200) <GET http://www.diningcity.com/netherlands/index.html> (res
[...]
```

DefaultHeadersMiddleware

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 289)

Unknown directive type "module".

```
.. module:: scrapy.downloadermiddlewares.defaultheaders
   :synopsis: Default Headers Downloader Middleware
```

This middleware sets all default requests headers specified in the `:setting:'DEFAULT_REQUEST_HEADERS'` setting.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 294);
[backlink](#)

Unknown interpreted text role "setting".

DownloadTimeoutMiddleware

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 300)

Unknown directive type "module".

```
.. module:: scrapy.downloadermiddlewares.downloadtimeout
   :synopsis: Download timeout middleware
```

This middleware sets the download timeout for requests specified in the `:setting:'DOWNLOAD_TIMEOUT'` setting or `:attr:'download_timeout'` spider attribute.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 305);

[backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 305);

[backlink](#)

Unknown interpreted text role "attr".

Note

You can also set download timeout per-request using `reqmeta:'download_timeout'` Request.meta key; this is supported even when DownloadTimeoutMiddleware is disabled.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 311); [backlink](#)

Unknown interpreted text role "reqmeta".

HttpAuthMiddleware

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 318)

Unknown directive type "module".

```
.. module:: scrapy.downloadermiddlewares.httputh
   :synopsis: HTTP Auth downloader middleware
```

This middleware authenticates all requests generated from certain spiders using [Basic access authentication](#) (aka. HTTP auth).

To enable HTTP authentication for a spider, set the `http_user` and `http_pass` spider attributes to the authentication data and the `http_auth_domain` spider attribute to the domain which requires this authentication (its subdomains will be also handled in the same way). You can set `http_auth_domain` to `None` to enable the authentication for all requests but you risk leaking your authentication credentials to unrelated domains.

Warning

In previous Scrapy versions HttpAuthMiddleware sent the authentication data with all requests, which is a security problem if the spider makes requests to several different domains. Currently if the `http_auth_domain` attribute is not set, the middleware will use the domain of the first request, which will work for some spiders but not for others. In the future the middleware will produce an error instead.

Example:

```
from scrapy.spiders import CrawlSpider

class SomeIntranetSiteSpider(CrawlSpider):

    http_user = 'someuser'
    http_pass = 'somepass'
    http_auth_domain = 'intranet.example.com'
    name = 'intranet.example.com'

    # .. rest of the spider code omitted ...
```

HttpCacheMiddleware

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 361)

Unknown directive type "module".

```
.. module:: scrapy.downloadermiddlewares.httpcache
   :synopsis: HTTP Cache downloader middleware
```

This middleware provides low-level cache to all HTTP requests and responses. It has to be combined with a cache storage backend as well as a cache policy.

Scrapy ships with the following HTTP cache storage backends:

- `ref:'httpcache-storage-fs'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 371); [backlink](#)

Unknown interpreted text role "ref".

- `ref:'httpcache-storage-dbm'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 372); [backlink](#)

Unknown interpreted text role "ref".

You can change the HTTP cache storage backend with the `'setting':HTTPCACHE_STORAGE` setting. Or you can also `ref:'implement your own storage backend. <httpcache-storage-custom>'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 374);

[backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 374);

[backlink](#)

Unknown interpreted text role "ref".

Scrapy ships with two HTTP cache policies:

- `.ref:'httpcache-policy-rfc2616'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 379); [backlink](#)

Unknown interpreted text role "ref".

- `.ref:'httpcache-policy-dummy'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 380); [backlink](#)

Unknown interpreted text role "ref".

You can change the HTTP cache policy with the `setting:'HTTPCACHE_POLICY'` setting. Or you can also implement your own policy.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 382); [backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 385)

Unknown directive type "reqmeta".

```
.. reqmeta:: dont_cache
```

You can also avoid caching a response on every policy using `reqmeta:'dont_cache'` meta key equals `True`.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 387); [backlink](#)

Unknown interpreted text role "reqmeta".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 389)

Unknown directive type "module".

```
.. module:: scrapy.extensions.httpcache
: noindex:
```

Dummy policy (default)

This policy has no awareness of any HTTP Cache-Control directives. Every request and its corresponding response are cached. When the same request is seen again, the response is returned without transferring anything from the Internet.

The Dummy policy is useful for testing spiders faster (without having to wait for downloads every time) and for trying your spider offline, when an Internet connection is not available. The goal is to be able to "replay" a spider run *exactly as it ran before*.

RFC2616 policy

This policy provides a RFC2616 compliant HTTP cache, i.e. with HTTP Cache-Control awareness, aimed at production and used in continuous runs to avoid downloading unmodified data (to save bandwidth and speed up crawls).

What is implemented:

- Do not attempt to store responses/requests with `no-store` cache-control directive set
- Do not serve responses from cache if `no-cache` cache-control directive is set even for fresh responses
- Compute freshness lifetime from `max-age` cache-control directive
- Compute freshness lifetime from `Expires` response header
- Compute freshness lifetime from `Last-Modified` response header (heuristic used by Firefox)
- Compute current age from `Age` response header
- Compute current age from `Date` header
- Revalidate stale responses based on `Last-Modified` response header
- Revalidate stale responses based on `ETag` response header
- Set `Date` header for any received response missing it
- Support `max-stale` cache-control directive in requests

This allows spiders to be configured with the full RFC2616 cache policy, but avoid revalidation on a request-by-request basis, while remaining conformant with the HTTP spec.

Example:

Add `Cache-Control: max-stale=600` to Request headers to accept responses that have exceeded their expiration time by no more than 600 seconds.

See also: RFC2616, 14.9.3

What is missing:

- Pragma: no-cache support <https://www.w3.org/Protocols/rfc2616/rfc2616-sec14.html#sec14.9.1>
- Vary header support <https://www.w3.org/Protocols/rfc2616/rfc2616-sec13.html#sec13.6>
- Invalidation after updates or deletes <https://www.w3.org/Protocols/rfc2616/rfc2616-sec13.html#sec13.10>
- ... probably others ..

Filesystem storage backend (default)

File system storage backend is available for the HTTP cache middleware.

Each request/response pair is stored in a different directory containing the following files:

- `request_body` - the plain request body
- `request_headers` - the request headers (in raw HTTP format)
- `response_body` - the plain response body
- `response_headers` - the request headers (in raw HTTP format)
- `meta` - some metadata of this cache resource in Python `repr()` format (grep-friendly format)
- `pickled_meta` - the same metadata in `meta` but pickled for more efficient deserialization

The directory name is made from the request fingerprint (see `scrapy.utils.request.fingerprint`), and one level of subdirectories is used to avoid creating too many files into the same directory (which is inefficient in many file systems). An example directory could be:

```
/path/to/cache/dir/example.com/72/72811f648e718090f041317756c03adb0ada46c7
```

DBM storage backend

A [DBM](#) storage backend is also available for the HTTP cache middleware.

By default, it uses the `'mod:dbm'`, but you can change it with the `setting:'HTTPCACHE_DBM_MODULE'` setting.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 497);
[backlink](#)

Unknown interpreted text role "mod".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 497);
[backlink](#)

Unknown interpreted text role "setting".

Writing your own storage backend

You can implement a cache storage backend by creating a Python class that defines the methods described below.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 508)

Unknown directive type "module".

```
.. module:: scrapy.extensions.httppcache
```

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 512)

Unknown directive type "method".

```
.. method:: open_spider(spider)
```

This method gets called after a spider has been opened for crawling. It handles the `:signal: open_spider <spider_opened>` signal.

```
:param spider: the spider which has been opened
:type spider: :class:`~scrapy.Spider` object
```

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 520)

Unknown directive type "method".

```
.. method:: close_spider(spider)
```

This method gets called after a spider has been closed. It handles the `:signal: close_spider <spider_closed>` signal.

```
:param spider: the spider which has been closed
:type spider: :class:`~scrapy.Spider` object
```

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 528)

Unknown directive type "method".

```
.. method:: retrieve_response(spider, request)
```

Return response if present in cache, or `None` otherwise.

```
:param spider: the spider which generated the request
:type spider: :class:`~scrapy.Spider` object
```

```
:param request: the request to find cached response for
:type request: :class:`~scrapy.Request` object
```

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 538)

Unknown directive type "method".

```
.. method:: store_response(spider, request, response)
```



```
Store the given response in the cache.

:param spider: the spider for which the response is intended
:type spider: :class:`~scrapy.Spider` object

:param request: the corresponding request the spider generated
:type request: :class:`~scrapy.Request` object

:param response: the response to store in the cache
:type response: :class:`~scrapy.http.Response` object
```

In order to use your storage backend, set:

- `setting: HTTPCACHE_STORAGE` to the Python import path of your custom storage class.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 553); [backlink](#)

Unknown interpreted text role "setting".

HTTPCache middleware settings

The `class:HttpCacheMiddleware` can be configured through the following settings:

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 559); [backlink](#)

Unknown interpreted text role "class".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 562)

Unknown directive type "setting".

```
.. setting:: HTTPCACHE_ENABLED
```

HTTPCACHE_ENABLED

Default: False

Whether the HTTP cache will be enabled.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 571)

Unknown directive type "setting".

```
.. setting:: HTTPCACHE_EXPIRATION_SECS
```

HTTPCACHE_EXPIRATION_SECS

Default: 0

Expiration time for cached requests, in seconds.

Cached requests older than this time will be re-downloaded. If zero, cached requests will never expire.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 583)

Unknown directive type "setting".

```
.. setting:: HTTPCACHE_DIR
```

HTTPCACHE_DIR

Default: 'httpcache'

The directory to use for storing the (low-level) HTTP cache. If empty, the HTTP cache will be disabled. If a relative path is given, is taken relative to the project data dir. For more info see: [ref:topics-project-structure](#).

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 590); [backlink](#)

Unknown interpreted text role "ref".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 594)

Unknown directive type "setting".

```
.. setting:: HTTPCACHE_IGNORE_HTTP_CODES
```

HTTPCACHE_IGNORE_HTTP_CODES

Default: []

Don't cache response with these HTTP codes.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 603)

Unknown directive type "setting".

```
.. setting:: HTTPCACHE_IGNORE_MISSING
```

HTTPCACHE_IGNORE_MISSING

Default: False

If enabled, requests not found in the cache will be ignored instead of downloaded.

```
System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 612)
```

Unknown directive type "setting".

```
.. setting:: HTTPCACHE_IGNORE_SCHEMES
```

HTTPCACHE_IGNORE_SCHEMES

Default: ['file']

Don't cache responses with these URI schemes.

```
System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 621)
```

Unknown directive type "setting".

```
.. setting:: HTTPCACHE_STORAGE
```

HTTPCACHE_STORAGE

Default: 'scrapy.extensions.httppache.FilesystemCacheStorage'

The class which implements the cache storage backend.

```
System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 630)
```

Unknown directive type "setting".

```
.. setting:: HTTPCACHE_DBM_MODULE
```

HTTPCACHE_DBM_MODULE

Default: 'dbm'

The database module to use in the [ref](#): DBM storage backend <httppache-storage-dbm>. This setting is specific to the DBM backend.

```
System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 637);
```

[backlink](#)

Unknown interpreted text role "ref".

```
System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 640)
```

Unknown directive type "setting".

```
.. setting:: HTTPCACHE_POLICY
```

HTTPCACHE_POLICY

Default: 'scrapy.extensions.httppache.DummyPolicy'

The class which implements the cache policy.

```
System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 649)
```

Unknown directive type "setting".

```
.. setting:: HTTPCACHE_GZIP
```

HTTPCACHE_GZIP

Default: False

If enabled, will compress all cached data with `gzip`. This setting is specific to the Filesystem backend.

```
System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 659)
```

Unknown directive type "setting".

```
.. setting:: HTTPCACHE_ALWAYS_STORE
```

HTTPCACHE_ALWAYS_STORE

Default: False

If enabled, will cache pages unconditionally.

A spider may wish to have all responses available in the cache, for future use with `Cache-Control: max-stale`, for instance. The `DummyPolicy` caches all responses but never revalidates them, and sometimes a more nuanced policy is desirable.

This setting still respects `Cache-Control: no-store` directives in responses. If you don't want that, filter `no-store` out of the `Cache-Control` headers in responses you feed to the cache middleware.

```
System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 677)
```

```
Unknown directive type "setting".
```

```
.. setting:: HTTPCACHE_IGNORE_RESPONSE_CACHE_CONTROLS
```

HTTPCACHE_IGNORE_RESPONSE_CACHE_CONTROLS

Default: []

List of Cache-Control directives in responses to be ignored.

Sites often set "no-store", "no-cache", "must-revalidate", etc., but get upset at the traffic a spider can generate if it actually respects those directives. This allows to selectively ignore Cache-Control directives that are known to be unimportant for the sites being crawled.

We assume that the spider will not issue Cache-Control directives in requests unless it actually needs them, so directives in requests are not filtered.

HttpCompressionMiddleware

```
System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 698)
```

```
Unknown directive type "module".
```

```
.. module:: scrapy.downloadermiddlewares.httpcompression
:synopsis: Http Compression Middleware
```

This middleware allows compressed (gzip, deflate) traffic to be sent/received from web sites.

This middleware also supports decoding [brotili-compressed](#) as well as [zstd-compressed](#) responses, provided that [brotili](#) or [zstandard](#) is installed, respectively.

HttpCompressionMiddleware Settings

```
System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 719)
```

```
Unknown directive type "setting".
```

```
.. setting:: COMPRESSION_ENABLED
```

COMPRESSION_ENABLED

Default: True

Whether the Compression middleware will be enabled.

HttpProxyMiddleware

```
System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 732)
```

```
Unknown directive type "module".
```

```
.. module:: scrapy.downloadermiddlewares.httpproxy
:synopsis: Http Proxy Middleware
```

```
System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 735)
```

```
Unknown directive type "reqmeta".
```

```
.. reqmeta:: proxy
```

This middleware sets the HTTP proxy to use for requests, by setting the `proxy` meta value for `class:~scrapy.Request` objects.

```
System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 739);
backlink
```

```
Unknown interpreted text role "class".
```

Like the Python standard library module `mod:urllib.request`, it obeys the following environment variables:

```
System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 742);
backlink
```

```
Unknown interpreted text role "mod".
```

- `http_proxy`
- `https_proxy`
- `no_proxy`

You can also set the meta key `proxy` per-request, to a value like `http://some_proxy_server:port` or `http://username:password@some_proxy_server:port`. Keep in mind this value will take precedence over `http_proxy/https_proxy` environment variables, and it will also ignore `no_proxy` environment variable.

RedirectMiddleware

```
System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 757)
```

```
Unknown directive type "module".
```

```
.. module:: scrapy.downloadermiddlewares.redirect
:synopsis: Redirection Middleware
```

This middleware handles redirection of requests based on response status.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 764)

Unknown directive type "reqmeta".

```
.. reqmeta:: redirect_urls
```

The urls which the request goes through (while being redirected) can be found in the `redirect_urls` attr: `Request.meta` <scrapy.Request.meta> key.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 766); [backlink](#)

Unknown interpreted text role "attr".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 769)

Unknown directive type "reqmeta".

```
.. reqmeta:: redirect_reasons
```

The reason behind each redirect in `reqmeta:redirect_urls` can be found in the `redirect_reasons` attr: `Request.meta` <scrapy.Request.meta> key. For example: [301, 302, 307, 'meta refresh'].

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 771); [backlink](#)

Unknown interpreted text role "reqmeta".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 771); [backlink](#)

Unknown interpreted text role "attr".

The format of a reason depends on the middleware that handled the corresponding redirect. For example, `:class:RedirectMiddleware` indicates the triggering response status code as an integer, while `:class:MetaRefreshMiddleware` always uses the 'meta refresh' string as reason.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 775); [backlink](#)

Unknown interpreted text role "class".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 775); [backlink](#)

Unknown interpreted text role "class".

The `:class:RedirectMiddleware` can be configured through the following settings (see the settings documentation for more info):

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 780); [backlink](#)

Unknown interpreted text role "class".

- `:setting:REDIRECT_ENABLED`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 783); [backlink](#)

Unknown interpreted text role "setting".

- `:setting:REDIRECT_MAX_TIMES`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 784); [backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 786)

Unknown directive type "reqmeta".

```
.. reqmeta:: dont_redirect
```

If `attr:Request.meta` <scrapy.Request.meta> has `dont_redirect` key set to True, the request will be ignored by this middleware.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 788); [backlink](#)

Unknown interpreted text role "attr".

If you want to handle some redirect status codes in your spider, you can specify these in the `handle_httpstatus_list` spider attribute.

For example, if you want the redirect middleware to ignore 301 and 302 responses (and pass them through to your spider) you can do this:

```
class MySpider(CrawlSpider):
    handle_httpstatus_list = [301, 302]
```

The `handle_httpstatus_list` key of `attr: Request.meta <scrapy.Request.meta>` can also be used to specify which response codes to allow on a per-request basis. You can also set the meta key `handle_httpstatus_all` to `True` if you want to allow any response code for a request.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 800); [backlink](#)

Unknown interpreted text role "attr".

RedirectMiddleware settings

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 810)

Unknown directive type "setting".

```
.. setting:: REDIRECT_ENABLED
```

REDIRECT_ENABLED

Default: True

Whether the Redirect middleware will be enabled.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 819)

Unknown directive type "setting".

```
.. setting:: REDIRECT_MAX_TIMES
```

REDIRECT_MAX_TIMES

Default: 20

The maximum number of redirections that will be followed for a single request. After this maximum, the request's response is returned as is.

MetaRefreshMiddleware

This middleware handles redirection of requests based on meta-refresh html tag.

The `:class:'MetaRefreshMiddleware'` can be configured through the following settings (see the settings documentation for more info):

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 836); [backlink](#)

Unknown interpreted text role "class".

- `:setting:'METAREFRESH_ENABLED'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 839); [backlink](#)

Unknown interpreted text role "setting".

- `:setting:'METAREFRESH_IGNORE_TAGS'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 840); [backlink](#)

Unknown interpreted text role "setting".

- `:setting:'METAREFRESH_MAXDELAY'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 841); [backlink](#)

Unknown interpreted text role "setting".

This middleware obey `:setting:'REDIRECT_MAX_TIMES'` setting, `:reqmeta:'dont_redirect'`, `:reqmeta:'redirect_urls'` and `:reqmeta:'redirect_reasons'` request meta keys as described for `:class:'RedirectMiddleware'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 843); [backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 843);
[backlink](#)

Unknown interpreted text role "reqmeta".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 843);
[backlink](#)

Unknown interpreted text role "reqmeta".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 843);
[backlink](#)

Unknown interpreted text role "reqmeta".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 843);
[backlink](#)

Unknown interpreted text role "class".

MetaRefreshMiddleware settings

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 851)

Unknown directive type "setting".

```
.. setting:: METAREFRESH_ENABLED
```

METAREFRESH_ENABLED

Default: True

Whether the Meta Refresh middleware will be enabled.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 860)

Unknown directive type "setting".

```
.. setting:: METAREFRESH_IGNORE_TAGS
```

METAREFRESH_IGNORE_TAGS

Default: []

Meta tags within these tags are ignored.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 869)

Unknown directive type "versionchanged".

```
.. versionchanged:: 2.0
   The default value of :setting:`METAREFRESH_IGNORE_TAGS` changed from
   ``['script', 'noscript']`` to ``[]``.
```

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 873)

Unknown directive type "setting".

```
.. setting:: METAREFRESH_MAXDELAY
```

METAREFRESH_MAXDELAY

Default: 100

The maximum meta-refresh delay (in seconds) to follow the redirection. Some sites use meta-refresh for redirecting to a session expired page, so we restrict automatic redirection to the maximum delay.

RetryMiddleware

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 887)

Unknown directive type "module".

```
.. module:: scrapy.downloadermiddlewares.retry
   :synopsis: Retry Middleware
```

A middleware to retry failed requests that are potentially caused by temporary problems such as a connection timeout or HTTP 500 error.

Failed pages are collected on the scraping process and rescheduled at the end, once the spider has finished crawling all regular (non failed) pages.

The `:class:'RetryMiddleware'` can be configured through the following settings (see the settings documentation for more info):

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 898); [backlink](#)

Unknown interpreted text role "class".

- `:setting:'RETRY_ENABLED'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 901); [backlink](#)

Unknown interpreted text role "setting".

- `:setting:'RETRY_TIMES'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 902); [backlink](#)

Unknown interpreted text role "setting".

- `:setting:'RETRY_HTTP_CODES'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 903); [backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 905)

Unknown directive type "reqmeta".

```
.. reqmeta:: dont_retry
```

If `:attr:'Request.meta <scrapy.Request.meta>'` has `dont_retry` key set to `True`, the request will be ignored by this middleware.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 907); [backlink](#)

Unknown interpreted text role "attr".

To retry requests from a spider callback, you can use the `:func:'get_retry_request'` function:

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 910); [backlink](#)

Unknown interpreted text role "func".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 913)

Unknown directive type "autofunction".

```
.. autofunction:: get_retry_request
```

RetryMiddleware Settings

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 918)

Unknown directive type "setting".

```
.. setting:: RETRY_ENABLED
```

RETRY_ENABLED

Default: `True`

Whether the Retry middleware will be enabled.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 927)

Unknown directive type "setting".

```
.. setting:: RETRY_TIMES
```

RETRY_TIMES

Default: `2`

Maximum number of times to retry, in addition to the first download.

Maximum number of retries can also be specified per-request using `reqmeta:'max_retry_times'` attribute of `:attr:'Request.meta <scrapy.Request.meta>'`. When initialized, the `reqmeta:'max_retry_times'` meta key takes higher precedence over the `:setting:'RETRY_TIMES'` setting.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 936);

[backlink](#)

Unknown interpreted text role "reqmeta".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 936);

[backlink](#)

Unknown interpreted text role "attr".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 936);

[backlink](#)

Unknown interpreted text role "reqmeta".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 936);

[backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 941)

Unknown directive type "setting".

```
.. setting:: RETRY_HTTP_CODES
```

RETRY_HTTP_CODES

Default: [500, 502, 503, 504, 522, 524, 408, 429]

Which HTTP response codes to retry. Other errors (DNS lookup issues, connections lost, etc) are always retried.

In some cases you may want to add 400 to `setting: 'RETRY_HTTP_CODES'` because it is a common code used to indicate server overload. It is not included by default because HTTP specs say so.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 951);

[backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 955)

Unknown directive type "setting".

```
.. setting:: RETRY_PRIORITY_ADJUST
```

RETRY_PRIORITY_ADJUST

Default: -1

Adjust retry request priority relative to original request:

- a positive priority adjust means higher priority.
- a negative priority adjust (default) means lower priority.

RobotsTxtMiddleware

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 973)

Unknown directive type "module".

```
.. module:: scrapy.downloadermiddlewares.robotstxt
:synopsis: robots.txt middleware
```

This middleware filters out requests forbidden by the robots.txt exclusion standard.

To make sure Scrapy respects robots.txt make sure the middleware is enabled and the `setting: 'ROBOTSTXT_OBEY'` setting is enabled.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 981);

[backlink](#)

Unknown interpreted text role "setting".

The `setting: 'ROBOTSTXT_USER_AGENT'` setting can be used to specify the user agent string to use for matching in the `robots.txt` file. If it is `None`, the User-Agent header you are sending with the request or the `setting: 'USER_AGENT'` setting (in that order) will be used for determining the user agent to use in the `robots.txt` file.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 984);

[backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 984);

[backlink](#)

Unknown interpreted text role "setting".

This middleware has to be combined with a [robots.txt](#) parser.

Scrapy ships with support for the following [robots.txt](#) parsers:

- `:ref: Protego <protego-parser>` (default)

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 994); [backlink](#)

Unknown interpreted text role "ref".

- `:ref: RobotFileParser <python-robotfileparser>`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 995); [backlink](#)

Unknown interpreted text role "ref".

- `:ref: Reppy <reppy-parser>`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 996); [backlink](#)

Unknown interpreted text role "ref".

- `:ref: Robotexclusionrulesparser <rrp-parser>`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 997); [backlink](#)

Unknown interpreted text role "ref".

You can change the [robots.txt](#) parser with the `:setting:ROBOTSTXT_PARSER` setting. Or you can also `:ref: implement support for a new parser <support-for-new-robots-parser>`.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 999); [backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 999); [backlink](#)

Unknown interpreted text role "ref".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1002)

Unknown directive type "reqmeta".

```
.. reqmeta:: dont_obey_robotstxt
```

If `:attr:Request.meta <scrapy.Request.meta>` has `dont_obey_robotstxt` key set to `True` the request will be ignored by this middleware even if `:setting:ROBOTSTXT_OBEY` is enabled.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1004); [backlink](#)

Unknown interpreted text role "attr".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1004); [backlink](#)

Unknown interpreted text role "setting".

Parsers vary in several aspects:

- Language of implementation
- Supported specification
- Support for wildcard matching
- Usage of [length based rule](#): in particular for `Allow` and `Disallow` directives, where the most specific rule based on the length of the path trumps the less specific (shorter) rule

Performance comparison of different parsers is available at [the following link](#).

Protego parser

Based on [Protego](#):

- implemented in Python
- is compliant with [Google's Robots.txt Specification](#)
- supports wildcard matching
- uses the length based rule

Scrapy uses this parser by default.

RobotFileParser

Based on `:class:`~urllib.robotparser.RobotFileParser``:

System Message: ERROR/3 (D: \onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1048); [backlink](#)

Unknown interpreted text role "class".

- is Python's built-in `robots.txt` parser
- is compliant with [Martijn Koster's 1996 draft specification](#)
- lacks support for wildcard matching
- doesn't use the length based rule

It is faster than Protego and backward-compatible with versions of Scrapy before 1.8.0.

In order to use this parser, set:

- `:setting`ROBOTSTXT_PARSER`` to `scrapy.robotstxt.PythonRobotParser`

System Message: ERROR/3 (D: \onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1063); [backlink](#)

Unknown interpreted text role "setting".

Reppy parser

Based on [Reppy](#):

- is a Python wrapper around [Robots Exclusion Protocol Parser for C++](#)
- is compliant with [Martijn Koster's 1996 draft specification](#)
- supports wildcard matching
- uses the length based rule

Native implementation, provides better speed than Protego.

In order to use this parser:

- Install [Reppy](#) by running `pip install reppy`

Warning

[Upstream issue #122](#) prevents reppy usage in Python 3.9+.

- Set `:setting`ROBOTSTXT_PARSER`` setting to `scrapy.robotstxt.ReppyRobotParser`

System Message: ERROR/3 (D: \onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1091); [backlink](#)

Unknown interpreted text role "setting".

Robotexclusionrulesparser

Based on [Robotexclusionrulesparser](#):

- implemented in Python
- is compliant with [Martijn Koster's 1996 draft specification](#)
- supports wildcard matching
- doesn't use the length based rule

In order to use this parser:

- Install [Robotexclusionrulesparser](#) by running `pip install robotexclusionrulesparser`
- Set `:setting`ROBOTSTXT_PARSER`` setting to `scrapy.robotstxt.RerpRobotParser`

System Message: ERROR/3 (D: \onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1116); [backlink](#)

Unknown interpreted text role "setting".

Implementing support for a new parser

You can implement support for a new `robots.txt` parser by subclassing the abstract base class `:class:`~scrapy.robotstxt.RobotParser`` and implementing the methods described below.

System Message: ERROR/3 (D: \onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1124); [backlink](#)

Unknown interpreted text role "class".

System Message: ERROR/3 (D: \onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1128)

Unknown directive type "module".

```
.. module:: scrapy.robotstxt
   :synopsis: robots.txt parser interface and implementations
```

System Message: ERROR/3 (D: \onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1131)

Unknown directive type "autoclass".

```
.. autoclass:: RobotParser
   :members:
```

DownloaderStats

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1139)

Unknown directive type "module".

```
.. module:: scrapy.downloadermiddlewares.stats
:synopsis: Downloader Stats Middleware
```

Middleware that stores stats of all requests, responses and exceptions that pass through it.

To use this middleware you must enable the `setting: DOWNLOADER_STATS` setting.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1147);
[backlink](#)

Unknown interpreted text role "setting".

UserAgentMiddleware

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1153)

Unknown directive type "module".

```
.. module:: scrapy.downloadermiddlewares.useragent
:synopsis: User Agent Middleware
```

Middleware that allows spiders to override the default user agent.

In order for a spider to override the default user agent, its `user_agent` attribute must be set.

AjaxCrawlMiddleware

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1168)

Unknown directive type "module".

```
.. module:: scrapy.downloadermiddlewares.ajaxcrawl
```

Middleware that finds 'AJAX crawlable' page variants based on meta-fragment html tag. See <https://developers.google.com/search/docs/ajax-crawling/docs/getting-started> for more info.

Note

Scrapy finds 'AJAX crawlable' pages for URLs like 'http://example.com/!#foo=bar' even without this middleware. AjaxCrawlMiddleware is necessary when URL doesn't contain '!#'. This is often a case for 'index' or 'main' website pages.

AjaxCrawlMiddleware Settings

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1187)

Unknown directive type "setting".

```
.. setting:: AJAXCRAWL_ENABLED
```

AJAXCRAWL_ENABLED

Default: False

Whether the AjaxCrawlMiddleware will be enabled. You may want to enable it for `ref: broad crawls <topics-broad-crawls>`.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1194);
[backlink](#)

Unknown interpreted text role "ref".

HttpProxyMiddleware settings

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1200)

Unknown directive type "setting".

```
.. setting:: HTTPPROXY_ENABLED
```

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1201)

Unknown directive type "setting".

```
.. setting:: HTTPPROXY_AUTH_ENCODING
```

HTTPPROXY_ENABLED

Default: True

Whether or not to enable the `:class:'HttpProxyMiddleware'`.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1208);
[backlink](#)

Unknown interpreted text role "class".

HTTPPROXY_AUTH_ENCODING

Default: "latin-1"

The default encoding for proxy authentication on `:class:'HttpProxyMiddleware'`.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\scrapy-master) (docs) (topics) downloader-middleware.rst, line 1215);
[backlink](#)

Unknown interpreted text role "class".