

Extensions

The extensions framework provides a mechanism for inserting your own custom functionality into Scrapy.

Extensions are just regular classes.

Extension settings

Extensions use the `:ref: Scrapy settings <topics-settings>` to manage their settings, just like any other Scrapy code.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 15); [backlink](#)

Unknown interpreted text role "ref".

It is customary for extensions to prefix their settings with their own name, to avoid collision with existing (and future) extensions. For example, a hypothetic extension to handle [Google Sitemaps](#) would use settings like `GOOGLESITEMAP_ENABLED`, `GOOGLESITEMAP_DEPTH`, and so on.

Loading & activating extensions

Extensions are loaded and activated at startup by instantiating a single instance of the extension class per spider being run. All the extension initialization code must be performed in the class `__init__` method.

To make an extension available, add it to the `:setting:'EXTENSIONS'` setting in your Scrapy settings. In `:setting:'EXTENSIONS'`, each extension is represented by a string: the full Python path to the extension's class name. For example:

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 32); [backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 32); [backlink](#)

Unknown interpreted text role "setting".

```
EXTENSIONS = {
    'scrapy.extensions.corestats.CoreStats': 500,
    'scrapy.extensions.telnet.TelnetConsole': 500,
}
```

As you can see, the `:setting:'EXTENSIONS'` setting is a dict where the keys are the extension paths, and their values are the orders, which define the extension *loading* order. The `:setting:'EXTENSIONS'` setting is merged with the `:setting:'EXTENSIONS_BASE'` setting defined in Scrapy (and not meant to be overridden) and then sorted by order to get the final sorted list of enabled extensions.

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

Unknown interpreted text role "setting".

As extensions typically do not depend on each other, their loading order is irrelevant in most cases. This is why the `:setting:'EXTENSIONS_BASE'` setting defines all extensions with the same order (0). However, this feature can be exploited if you need to add an extension which depends on other extensions already loaded.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 49); [backlink](#)

Unknown interpreted text role "setting".

Available, enabled and disabled extensions

Not all available extensions will be enabled. Some of them usually depend on a particular setting. For example, the HTTP Cache extension is available by default but disabled unless the `:setting:'HTTPCACHE_ENABLED'` setting is set.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 58); [backlink](#)

Unknown interpreted text role "setting".

Disabling an extension

In order to disable an extension that comes enabled by default (i.e. those included in the `:setting:'EXTENSIONS_BASE'` setting) you must set its order to `None`. For example:

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 65); [backlink](#)

Unknown interpreted text role "setting".

```
EXTENSIONS = {
    'scrapy.extensions.corestats.CoreStats': None,
}
```

Writing your own extension

Each extension is a Python class. The main entry point for a Scrapy extension (this also includes middlewares and pipelines) is the `from_crawler` class method which receives a `Crawler` instance. Through the `Crawler` object you can access settings, signals, stats, and also control the crawling behaviour.

Typically, extensions connect to `:ref:signals <topics-signals>` and perform tasks triggered by them.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 81); [backlink](#)

Unknown interpreted text role "ref".

Finally, if the `from_crawler` method raises the `:exc:`~scrapy.exceptions.NotConfigured`` exception, the extension will be disabled. Otherwise, the extension will be enabled.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 84); [backlink](#)

Unknown interpreted text role "exc".

Sample extension

Here we will implement a simple extension to illustrate the concepts described in the previous section. This extension will log a message every time:

- a spider is opened
- a spider is closed
- a specific number of items are scraped

The extension will be enabled through the `MYEXT_ENABLED` setting and the number of items will be specified through the `MYEXT_ITEMCOUNT` setting.

Here is the code of such extension:

```
import logging
from scrapy import signals
from scrapy.exceptions import NotConfigured

logger = logging.getLogger(__name__)
```

```

class SpiderOpenCloseLogging:

    def __init__(self, item_count):
        self.item_count = item_count
        self.items_scraped = 0

    @classmethod
    def from_crawler(cls, crawler):
        # first check if the extension should be enabled and raise
        # NotConfigured otherwise
        if not crawler.settings.getbool('MYEXT_ENABLED'):
            raise NotConfigured

        # get the number of items from settings
        item_count = crawler.settings.getint('MYEXT_ITEMCOUNT', 1000)

        # instantiate the extension object
        ext = cls(item_count)

        # connect the extension object to signals
        crawler.signals.connect(ext.spider_opened, signal=signals.spider_opened)
        crawler.signals.connect(ext.spider_closed, signal=signals.spider_closed)
        crawler.signals.connect(ext.item_scraped, signal=signals.item_scraped)

        # return the extension object
        return ext

    def spider_opened(self, spider):
        logger.info("opened spider %s", spider.name)

    def spider_closed(self, spider):
        logger.info("closed spider %s", spider.name)

    def item_scraped(self, item, spider):
        self.items_scraped += 1
        if self.items_scraped % self.item_count == 0:
            logger.info("scraped %d items", self.items_scraped)

```

Built-in extensions reference

General purpose extensions

Log Stats extension

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 159)

Unknown directive type "module".

```

.. module:: scrapy.extensions.logstats
   :synopsis: Basic stats logging

```

Log basic stats like crawled pages and scraped items.

Core Stats extension

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 169)

Unknown directive type "module".

```

.. module:: scrapy.extensions.corestats
   :synopsis: Core stats collection

```

Enable the collection of core statistics, provided the stats collection is enabled (see [ref:topics-stats](#)).

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 174); [backlink](#)

Unknown interpreted text role "ref".

Telnet console extension

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

Unknown directive type "module".

```
.. module:: scrapy.extensions.telnet
   :synopsis: Telnet console
```

Provides a telnet console for getting into a Python interpreter inside the currently running Scrapy process, which can be very useful for debugging.

The telnet console must be enabled by the `:setting:'TELNETCONSOLE_ENABLED'` setting, and the server will listen in the port specified in `:setting:'TELNETCONSOLE_PORT'`.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 190); [backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 190); [backlink](#)

Unknown interpreted text role "setting".

Memory usage extension

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 199)

Unknown directive type "module".

```
.. module:: scrapy.extensions.memusage
   :synopsis: Memory usage extension
```

Note

This extension does not work in Windows.

Monitors the memory used by the Scrapy process that runs the spider and:

1. sends a notification e-mail when it exceeds a certain value
2. closes the spider when it exceeds a certain value

The notification e-mails can be triggered when a certain warning value is reached (`:setting:'MEMUSAGE_WARNING_MB'`) and when the maximum value is reached (`:setting:'MEMUSAGE_LIMIT_MB'`) which will also cause the spider to be closed and the Scrapy process to be terminated.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 211); [backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 211); [backlink](#)

Unknown interpreted text role "setting".

This extension is enabled by the `:setting:'MEMUSAGE_ENABLED'` setting and can be configured with the following settings:

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 216); [backlink](#)

Unknown interpreted text role "setting".

- `:setting:'MEMUSAGE_LIMIT_MB'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 219); [backlink](#)

Unknown interpreted text role "setting".

- :setting:`MEMUSAGE_WARNING_MB`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 220); [backlink](#)

Unknown interpreted text role "setting".

- :setting:`MEMUSAGE_NOTIFY_MAIL`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 221); [backlink](#)

Unknown interpreted text role "setting".

- :setting:`MEMUSAGE_CHECK_INTERVAL_SECONDS`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 222); [backlink](#)

Unknown interpreted text role "setting".

Memory debugger extension

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 227)

Unknown directive type "module".

```
.. module:: scrapy.extensions.memdebug
   :synopsis: Memory debugger extension
```

An extension for debugging memory usage. It collects information about:

- objects uncollected by the Python garbage collector
- objects left alive that shouldn't. For more info, see :ref:`topics-leaks-trackrefs`

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

Unknown interpreted text role "ref".

To enable this extension, turn on the :setting:`MEMDEBUG_ENABLED` setting. The info will be stored in the stats.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 237); [backlink](#)

Unknown interpreted text role "setting".

Close spider extension

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 243)

Unknown directive type "module".

```
.. module:: scrapy.extensions.closespider
```

:synopsis: Close spider extension

Closes a spider automatically when some conditions are met, using a specific closing reason for each condition.

The conditions for closing a spider can be configured through the following settings:

- `:setting:'CLOSESPIDER_TIMEOUT'`

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

Unknown interpreted text role "setting".

- `:setting:'CLOSESPIDER_ITEMCOUNT'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 255); [backlink](#)

Unknown interpreted text role "setting".

- `:setting:'CLOSESPIDER_PAGECOUNT'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 256); [backlink](#)

Unknown interpreted text role "setting".

- `:setting:'CLOSESPIDER_ERRORCOUNT'`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 257); [backlink](#)

Unknown interpreted text role "setting".

Note

When a certain closing condition is met, requests which are currently in the downloader queue (up to `:setting:'CONCURRENT_REQUESTS'` requests) are still processed.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 261); [backlink](#)

Unknown interpreted text role "setting".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 265)

Unknown directive type "setting".

```
.. setting:: CLOSESPIDER_TIMEOUT
```

CLOSESPIDER_TIMEOUT

Default: 0

An integer which specifies a number of seconds. If the spider remains open for more than that number of second, it will be automatically closed with the reason `closespider_timeout`. If zero (or non set), spiders won't be closed by timeout.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 277)

Unknown directive type "setting".

```
.. setting:: CLOSESPIDER_ITEMCOUNT
```

CLOSESPIDER_ITEMCOUNT

Default: 0

An integer which specifies a number of items. If the spider scrapes more than that amount and those items are passed by the item pipeline, the spider will be closed with the reason `close_spider_itemcount`. If zero (or non set), spiders won't be closed by number of passed items.

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

Unknown directive type "setting".

```
.. setting:: CLOSESPIDER_PAGECOUNT
```

CLOSESPIDER_PAGECOUNT

Default: 0

An integer which specifies the maximum number of responses to crawl. If the spider crawls more than that, the spider will be closed with the reason `close_spider_pagecount`. If zero (or non set), spiders won't be closed by number of crawled responses.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 301)

Unknown directive type "setting".

```
.. setting:: CLOSESPIDER_ERRORCOUNT
```

CLOSESPIDER_ERRORCOUNT

Default: 0

An integer which specifies the maximum number of errors to receive before closing the spider. If the spider generates more than that number of errors, it will be closed with the reason `close_spider_errorcount`. If zero (or non set), spiders won't be closed by number of errors.

StatsMailer extension

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 316)

Unknown directive type "module".

```
.. module:: scrapy.extensions.statsmailer
   :synopsis: StatsMailer extension
```

This simple extension can be used to send a notification e-mail every time a domain has finished scraping, including the Scrapy stats collected. The email will be sent to all recipients specified in the `:setting:'STATSMAILER_RCPTS'` setting.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 321); [backlink](#)

Unknown interpreted text role "setting".

Emails can be sent using the `:class:'~scrapy.mail.MailSender'` class. To see a full list of parameters, including examples on how to instantiate `:class:'~scrapy.mail.MailSender'` and use mail settings, see `ref:'topics-email'`.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 326); [backlink](#)

Unknown interpreted text role "class".

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

master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 326); [backlink](#)

Unknown interpreted text role "class".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 326); [backlink](#)

Unknown interpreted text role "ref".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 331)

Unknown directive type "module".

```
.. module:: scrapy.extensions.debug
   :synopsis: Extensions for debugging Scrapy
```

Debugging extensions

Stack trace dump extension

Dumps information about the running process when a [SIGQUIT](#) or [SIGUSR2](#) signal is received. The information dumped is the following:

1. engine status (using `scrapy.utils.engine.get_engine_status()`)
2. live references (see [ref`topics-leaks-trackrefs`](#))

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\topics\[scrapy-master] [docs] [topics]extensions.rst, line 346); [backlink](#)

Unknown interpreted text role "ref".

3. stack trace of all threads

After the stack trace and engine status is dumped, the Scrapy process continues running normally.

This extension only works on POSIX-compliant platforms (i.e. not Windows), because the [SIGQUIT](#) and [SIGUSR2](#) signals are not available on Windows.

There are at least two ways to send Scrapy the [SIGQUIT](#) signal:

1. By pressing Ctrl-while a Scrapy process is running (Linux only?)
2. By running this command (assuming `<pid>` is the process id of the Scrapy process):

```
kill -QUIT <pid>
```

Debugger extension

Invokes a `.doc:Python debugger <library/pdb>` inside a running Scrapy process when a [SIGUSR2](#) signal is received. After the debugger is exited, the Scrapy process continues running normally.

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

Unknown interpreted text role "doc".

For more info see [Debugging in Python](#).

This extension only works on POSIX-compliant platforms (i.e. not Windows).