
Scrapy Pipelines

Scrapedia

Jun 12, 2019

CONTENTS:

1	Installation	1
2	ItemPipeline	3
3	Pipeline MongoDB	5
4	Items	7
5	Settings	9
6	Signals	11
7	Indices and tables	13
	Python Module Index	15
	Index	17

INSTALLATION

ITEMPIPELINE

```
class scrapy_pipelines.pipelines.ItemPipeline (settings: scrapy.settings.Settings = None)
    Abstract Class for the item pipeline
```

```
    abstract close_spider (spider: scrapy.spiders.Spider)
```

Parameters **spider** (*Spider*) –

Returns

Return type

```
    classmethod from_crawler (crawler: scrapy.crawler.Crawler)
```

Parameters **crawler** (*Crawler*) –

Returns

Return type

```
    abstract classmethod from_settings (settings: scrapy.settings.Settings)
```

Parameters **settings** (*Settings*) –

Returns

Return type

```
    abstract open_spider (spider: scrapy.spiders.Spider)
```

Parameters **spider** (*Spider*) –

Returns

Return type

```
    abstract process_item (item: scrapy.item.Item, spider: scrapy.spiders.Spider) →
                           scrapy.item.Item
```

Parameters

- **item** (*Item*) –
- **spider** (*Spider*) –

Returns

Return type Item

PIPELINE MONGODB

class scrapy_pipelines.pipelines.mongo.**MongoPipeline** (*uri: str, settings: scrapy.settings.Settings*)

A pipeline saved into MongoDB asynchronously with txmongo

close_spider (*spider: scrapy.spiders.Spider*)

Parameters **spider** (*Spider*) –

Returns

Return type

create_indexes (*spider: scrapy.spiders.Spider*)

Parameters **spider** (*Spider*) –

Returns

Return type

classmethod from_crawler (*crawler: scrapy.crawler.Crawler*)

Parameters **crawler** (*Crawler*) –

Returns

Return type *MongoPipeline*

classmethod from_settings (*settings: scrapy.settings.Settings*)

Parameters **settings** (*Settings*) –

Returns

Return type *MongoPipeline*

item_completed (*result: str, item: scrapy.item.Item, spider: scrapy.spiders.Spider*) → scrapy.item.Item

Parameters

- **result** (*str*) –
- **item** (*Item*) –
- **spider** (*Spider*) –

Returns

Return type Item

open_spider (*spider: scrapy.spiders.Spider*)

Parameters **spider** (*Spider*) –

Returns

Return type

process_item (*item*: scrapy.item.Item, *spider*: scrapy.spiders.Spider) → scrapy.item.Item

Parameters

- **item** (*Item*) –
- **spider** (*Spider*) –

Returns

Return type Item

process_item_id (*signal*: object, *sender*: scrapy.crawler.Crawler, *item*: scrapy.item.Item, *spider*: scrapy.spiders.Spider) → pymongo.results.InsertOneResult

Parameters

- **signal** (*object*) –
- **sender** (*Crawler*) –
- **item** (*Item*) –
- **spider** (*Spider*) –

Returns

Return type InsertOneResult

A customized item for MongoDB

```
class scrapy_pipelines.items.BSONItem(*args, **kwargs)
    Pymongo creates _id automatically in the object after inserting
```


SETTINGS

The utilities used in settings module

`scrapy_pipelines.settings.unfreeze_settings` (*settings*: `scrapy.settings.Settings`) → Generator[`scrapy.settings.Settings`, None, None]

Parameters `settings` (*Settings*) –

Returns

Return type Generator[Settings, None, None]

SIGNALS

Signals for the pipelines

Installation Installation

ItemPipeline The root class for all pipelines

Pipeline MongoDB Save items into MongoDB

Items Items used in these pipelines

Settings Settings for these pipelines

Signals Signals used in these pipelines

INDICES AND TABLES

- `genindex`
- `modindex`
- `search`

PYTHON MODULE INDEX

S

`scrapy_pipelines.items`, [7](#)
`scrapy_pipelines.settings`, [9](#)
`scrapy_pipelines.signals`, [11](#)

INDEX

B

BSONItem (class in scrapy_pipelines.items), 7

C

close_spider() (scrapy_pipelines.pipelines.ItemPipeline
method), 3

close_spider() (scrapy_pipelines.pipelines.mongo.MongoPipeline
method), 5

create_indexes() (scrapy_pipelines.pipelines.mongo.MongoPipeline
method), 5

F

from_crawler() (scrapy_pipelines.pipelines.ItemPipeline
class method), 3

from_crawler() (scrapy_pipelines.pipelines.mongo.MongoPipeline
class method), 5

from_settings() (scrapy_pipelines.pipelines.ItemPipeline
class method), 3

from_settings() (scrapy_pipelines.pipelines.mongo.MongoPipeline
class method), 5

I

item_completed() (scrapy_pipelines.pipelines.mongo.MongoPipeline
method), 5

ItemPipeline (class in scrapy_pipelines.pipelines), 3

M

MongoPipeline (class in scrapy_pipelines.pipelines.mongo), 5

O

open_spider() (scrapy_pipelines.pipelines.ItemPipeline
method), 3

open_spider() (scrapy_pipelines.pipelines.mongo.MongoPipeline
method), 5

P

process_item() (scrapy_pipelines.pipelines.ItemPipeline
method), 3

process_item() (scrapy_pipelines.pipelines.mongo.MongoPipeline
method), 6

process_item_id()

(scrapy_pipelines.pipelines.mongo.MongoPipeline
method), 6

S

scrapy_pipelines.items (module), 7

scrapy_pipelines.settings (module), 9

scrapy_pipelines.signals (module), 11

U

unfreeze_settings() (in scrapy_pipelines.settings), 9 module