

Booklook Crawlers

0.9

Generated by Doxygen 1.8.9.1

Thu Apr 23 2015 14:45:32

Contents

Chapter 1

Hierarchical Index

1.1 Class Hierarchy

This inheritance list is sorted roughly, but not completely, alphabetically:

| | |
|---|----|
| bookcrawl.assigner.BookAssigner | ?? |
| Item | |
| bookcrawl.items.AmazonPageItem | ?? |
| bookcrawl.items.AmazonSearchItem | ?? |
| bookcrawl.items.BookcrawlItem | ?? |
| object | |
| bookcrawl.pipelines.BookcrawlPipeline | ?? |

Chapter 2

Class Index

2.1 Class List

Here are the classes, structs, unions and interfaces with brief descriptions:

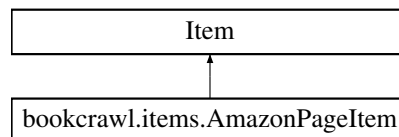
| | |
|---|----|
| bookcrawl.items.AmazonPageItem | ?? |
| bookcrawl.items.AmazonSearchItem | ?? |
| bookcrawl.assigner.BookAssigner | ?? |
| bookcrawl.items.BookcrawlItem | ?? |
| bookcrawl.pipelines.BookcrawlPipeline | ?? |

Chapter 3

Class Documentation

3.1 bookcrawl.items.AmazonPageItem Class Reference

Inheritance diagram for bookcrawl.items.AmazonPageItem:



Static Public Attributes

- tuple **p_title** = scrapy.Field()
- tuple **p_author** = scrapy.Field()
- tuple **p_pages** = scrapy.Field()

3.1.1 Detailed Description

Part of a trio of objects designed to have data passed to by the appropriate spider. Their fields are flushed to a JSON-formatted file when their jobs are complete.

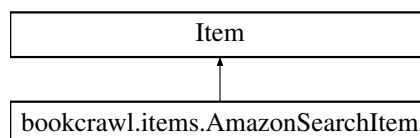
This item is associated with the BookSpider.

The documentation for this class was generated from the following file:

- items.py

3.2 bookcrawl.items.AmazonSearchItem Class Reference

Inheritance diagram for bookcrawl.items.AmazonSearchItem:



Static Public Attributes

- tuple **s_title** = scrapy.Field()
- tuple **s_url** = scrapy.Field()

3.2.1 Detailed Description

Part of a trio of objects designed to have data passed to by the appropriate spider. Their fields are flushed to a JSON-formatted file when their jobs are complete.

This item is associated with the AmazonSpider.

The documentation for this class was generated from the following file:

- items.py

3.3 bookcrawl.assigner.BookAssigner Class Reference

Public Member Functions

- def **__init__** (self, path_to_page_file, path_to_auth_title_file)
- def **assignment_iterator** (self)

Public Attributes

- **page_path**
- **title_path**

Static Public Attributes

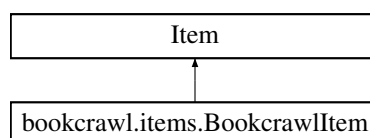
- string **page_path** = ""
- string **title_path** = ""
- **assign_file** = None
- tuple **assignments** = Queue.Queue()
- tuple **jsondata** = json.load(self.assign_file)

The documentation for this class was generated from the following file:

- assigner.py

3.4 bookcrawl.items.BookcrawlItem Class Reference

Inheritance diagram for bookcrawl.items.BookcrawlItem:



Static Public Attributes

- tuple **title** = scrapy.Field()
- tuple **url** = scrapy.Field()
- tuple **pages** = scrapy.Field()
- tuple **author** = scrapy.Field()

3.4.1 Detailed Description

Part of a trio of objects designed to have data passed to by the appropriate spider. Their fields are flushed to a JSON-formatted file when their jobs are complete.

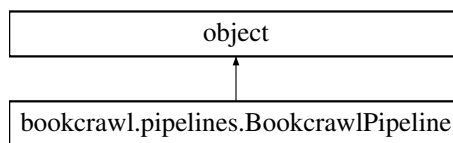
This item is associated with the BookSpider.

The documentation for this class was generated from the following file:

- items.py

3.5 bookcrawl.pipelines.BookcrawlPipeline Class Reference

Inheritance diagram for bookcrawl.pipelines.BookcrawlPipeline:



Public Member Functions

- def **__init__** (self)
- def **spider_opened** (self, spider)
- def **spider_closed** (self, spider)
- def **process_item** (self, item, spider)

Public Attributes

- **files**
- **exporter**

The documentation for this class was generated from the following file:

- pipelines.py

