



Test Automation

Scraping Images through scrapy

Submitted by :

Rahul Gusain

Submitted to :

Mr. Alind A

```

(exp3) Rahuls-Air:~ rahulgusain$ scrapy startproject imageextr
New Scrapy project 'imageextr', using template directory '/opt/homebrew/Caskroom/miniconda/base/envs/exp3/lib/python3.10/site-packages/scrapy/templates/project', created in:
/Users/rahulgusain/imageextr

You can start your first spider with:
cd imageextr
scrapy genspider example example.com
(exp3) Rahuls-Air:~ rahulgusain$ cd imageextr
(exp3) Rahuls-Air:imageextr rahulgusain$ ls
imageextr  scrapy.cfg
(exp3) Rahuls-Air:imageextr rahulgusain$ cd imageextr
(exp3) Rahuls-Air:imageextr rahulgusain$ ls
__init__.py  items.py  middlewares.py  pipelines.py  settings.py  spiders
(exp3) Rahuls-Air:imageextr rahulgusain$ cd spiders
(exp3) Rahuls-Air:spiders rahulgusain$ ls
__init__.py
(exp3) Rahuls-Air:spiders rahulgusain$ scrapy genspider image image.com
Created spider 'image' using template 'basic' in module:
imageextr.spiders.image
(exp3) Rahuls-Air:spiders rahulgusain$ ls
__init__.py  __pycache__  image.py
(exp3) Rahuls-Air:spiders rahulgusain$

```

```

Two_Sum.py  sqrt.py  image.py  Release Notes: 1.75.0
Users > rahulgusain > imageextr > imageextr > spiders > image.py
1  import requests
2
3
4
5  def imgs():
6      with open("image.jpg", "wb") as f:
7          for chunk in response.iter_content(chunk_size=512):
8              f.write(chunk)
9
10
11
12  class ImageSpider(scrapy.Spider):
13      name = "image_spider"
14      start_urls = [
15          'https://www.freepik.com/free-photos-vectors/nature-wallpaper',
16      ]
17
18      def parse(self, response):
19          for image in response.css('img'):
20              yield {
21                  'image_url': image.css('img::attr(src)').get()
22              }
23
24
25

```

VS Code interface showing the file explorer with 'Two_Sum.py', 'sqrt.py', and 'image.py'. The bottom status bar indicates 'Ln 18, Col 42', 'Spaces: 4', 'UTF-8', 'LF', 'Python', and 'Go Live'.

```
{'image_url': 'https://avatar.freepik.com/18831218.jpg'}
2023-02-10 12:24:16 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-wallpaper>
{'image_url': 'https://freepik.cdnpk.net/img/logos/freepikcompany-footer.svg'}
2023-02-10 12:24:16 [scrapy.core.engine] INFO: Closing spider (finished)
2023-02-10 12:24:16 [scrapy.extensions.feedexport] INFO: Stored csv feed (153 items) in: test.csv
2023-02-10 12:24:16 [scrapy.statscollectors] INFO: Dumping Scrapy stats:
{'downloader/request_bytes': 484,
 'downloader/request_count': 2,
 'downloader/request_method_count/GET': 2,
 'downloader/response_bytes': 44287,
 'downloader/response_count': 2,
 'downloader/response_status_count/200': 2,
 'elapsed_time_seconds': 2.184223,
 'feedexport/success_count/FileFeedStorage': 1,
 'finish_reason': 'finished',
 'finish_time': datetime.datetime(2023, 2, 10, 6, 54, 16, 813331),
 'httpcompression/response_bytes': 313321,
 'httpcompression/response_count': 1,
 'item_scraped_count': 153,
 'log_count/DEBUG': 166,
 'log_count/INFO': 11,
 'memusage/max': 69795840,
 'memusage/startup': 69779456,
 'response_received_count': 2,
 'robotstxt/request_count': 1,
 'robotstxt/response_count': 1,
 'robotstxt/response_status_count/200': 1,
 'scheduler/dequeued': 1,
 'scheduler/dequeued/memory': 1,
 'scheduler/enqueued': 1,
 'scheduler/enqueued/memory': 1,
 'start_time': datetime.datetime(2023, 2, 10, 6, 54, 14, 629108)}
2023-02-10 12:24:16 [scrapy.core.engine] INFO: Spider closed (finished)
(exp3) Rahul's-Air:spiders rahulgusain$
```

```
in _evaluate_iterable
  for r in iterable:
    File "/Users/rahulgusain/imageextr/imageextr/spiders/image.py", line 21, in parse
      urllib.request.urlretrieve(image_url, image_url.split('/')[1])
NameError: name 'urllib' is not defined
2023-02-10 12:26:18 [scrapy.core.engine] INFO: Closing spider (finished)
2023-02-10 12:26:18 [scrapy.statscollectors] INFO: Dumping Scrapy stats:
{'downloader/request_bytes': 484,
 'downloader/request_count': 2,
 'downloader/request_method_count/GET': 2,
 'downloader/response_bytes': 44275,
 'downloader/response_count': 2,
 'downloader/response_status_count/200': 2,
 'elapsed_time_seconds': 1.03805,
 'finish_reason': 'finished',
 'finish_time': datetime.datetime(2023, 2, 10, 6, 56, 18, 926284),
 'httpcompression/response_bytes': 313321,
 'httpcompression/response_count': 1,
 'log_count/DEBUG': 11,
 'log_count/ERROR': 1,
 'log_count/INFO': 10,
 'memusage/max': 69697536,
 'memusage/startup': 69681152,
 'response_received_count': 2,
 'robotstxt/request_count': 1,
 'robotstxt/response_count': 1,
 'robotstxt/response_status_count/200': 1,
 'scheduler/dequeued': 1,
 'scheduler/dequeued/memory': 1,
 'scheduler/enqueued': 1,
 'scheduler/enqueued/memory': 1,
 'spider_exceptions/NameError': 1,
 'start_time': datetime.datetime(2023, 2, 10, 6, 56, 17, 888234)}
2023-02-10 12:26:18 [scrapy.core.engine] INFO: Spider closed (finished)
(exp3) Rahul's-Air:spiders rahulgusain$
```



```
allpaper>
{'image_url': 'https://avatar.cdnpk.net/default_01.png'}
2023-02-10 12:29:45 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://freepik.cdnpk.net/img/1px.png'}
2023-02-10 12:29:45 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://img.freepik.com/premium-photo/beautiful-landscape-based-3d-rendering-illustration_771975-25.jpg'
}
2023-02-10 12:29:49 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://lh3.googleusercontent.com/a/ALm5wu2MVIBN_odVqk0nWb4yAuTUmH3nw_NyTmmak0p=s96-c'}
2023-02-10 12:29:49 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://freepik.cdnpk.net/img/1px.png'}
2023-02-10 12:29:50 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://img.freepik.com/premium-photo/beautiful-moraine-lake-banff-national-park-alberta-canada_131985-9
8.jpg'}
2023-02-10 12:29:50 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://lh3.googleusercontent.com/-0EQovBH-wVs/AAAAAAAAAI/AAAAAAAAADQ/cue0JIaSrKc/photo.jpg?sz=250'}
2023-02-10 12:29:50 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://freepik.cdnpk.net/img/1px.png'}
2023-02-10 12:29:51 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://img.freepik.com/free-photo/waterfall-clean-tourist-blue-flow-asian_1417-1350.jpg'}
2023-02-10 12:29:51 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://avatar.cdnpk.net/default_01.png'}
2023-02-10 12:29:52 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://freepik.cdnpk.net/img/1px.png'}
```



```
und_181624-12847.jpg'}
2023-02-10 12:29:55 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://avatar.freepik.com/18831218.jpg'}
2023-02-10 12:29:55 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://freepik.cdnpk.net/img/1px.png'}
2023-02-10 12:29:56 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://img.freepik.com/free-photo/vertical-orange-lily-field-cloudy-dark-sky_181624-37905.jpg'}
2023-02-10 12:29:57 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://avatar.freepik.com/18831218.jpg'}
2023-02-10 12:29:57 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://freepik.cdnpk.net/img/1px.png'}
2023-02-10 12:29:58 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://img.freepik.com/free-photo/lone-tree_181624-46361.jpg'}
2023-02-10 12:29:58 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://avatar.freepik.com/18831218.jpg'}
2023-02-10 12:29:58 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://freepik.cdnpk.net/img/1px.png'}
2023-02-10 12:29:59 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://img.freepik.com/free-photo/beautiful-scenery-pathway-forest-with-trees-covered-with-frost_181624
-42376.jpg'}
2023-02-10 12:29:59 [scrapy.core.scrapers] DEBUG: Scraped from <200 https://www.freepik.com/free-photos-vectors/nature-w
allpaper>
{'image_url': 'https://avatar.freepik.com/18831218.jpg'}
^Z
[1]+  Stopped                  scrapy crawl image_spider -o test.csv
(exp3) Rahul's-Air:spiders rahulgusain$
```



