



Test Automation

Experiment 3

Submitted by :

Rahul Gusain

Submitted to :

Mr. Alind A

Step 1: - Write the below commands in the terminal or paste the below commands in the terminal to create the spyder directory. It will contain some more files as shown in screenshot which is automatically being created by the scrapy.

Command: - **scrapy startproject spyder**

```
(base) Rahuls-Air:~ rahulgusain$ conda create --name exp3
Collecting package metadata (current_repodata.json): done
Solving environment: done

==> WARNING: A newer version of conda exists. <==
  current version: 4.12.0
  latest version: 23.1.0

Please update conda by running

    $ conda update -n base -c defaults conda

## Package Plan ##

  environment location: /opt/homebrew/Caskroom/miniconda/base/envs/exp3

Proceed ([y]/n)? y
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
#
# To activate this environment, use
#
#     $ conda activate exp3
#
# To deactivate an active environment, use
#
#     $ conda deactivate
```

```

Please update conda by running

$ conda update -n base -c defaults conda

## Package Plan ##

environment location: /opt/homebrew/Caskroom/miniconda/base/envs/exp3

Proceed ([y]/n)? y

Preparing transaction: done
Verifying transaction: done
Executing transaction: done
#
# To activate this environment, use
#
# $ conda activate exp3
#
# To deactivate an active environment, use
#
# $ conda deactivate

(base) Rahuls-Air:~ rahulgusain$ conda env list
# conda environments:
#
base                * /opt/homebrew/Caskroom/miniconda/base
exp1                 /opt/homebrew/Caskroom/miniconda/base/envs/exp1
exp3                 /opt/homebrew/Caskroom/miniconda/base/envs/exp3

(base) Rahuls-Air:~ rahulgusain$ conda activate exp3
(exp3) Rahuls-Air:~ rahulgusain$

```

```

requests            pkgs/main/osx-arm64::requests-2.28.1-py310hca03da5_0
requests-file       pkgs/main/noarch::requests-file-1.5.1-pyhd3eb1b0_0
scrappy             pkgs/main/osx-arm64::scrappy-2.6.2-py310hca03da5_0
service_identity    pkgs/main/noarch::service_identity-18.1.0-pyhd3eb1b0_1
setuptools          pkgs/main/osx-arm64::setuptools-65.6.3-py310hca03da5_0
six                pkgs/main/noarch::six-1.16.0-pyhd3eb1b0_1
sqlite             pkgs/main/osx-arm64::sqlite-3.40.1-h7a7dc30_0
tk                 pkgs/main/osx-arm64::tk-8.6.12-hb8d0fd4_0
tldextract          pkgs/main/noarch::tldextract-3.2.0-pyhd3eb1b0_0
twisted            pkgs/main/osx-arm64::twisted-22.2.0-py310h1a28f6b_1
typing_extensions   pkgs/main/osx-arm64::typing_extensions-4.4.0-py310hca03da5_0
tzdata             pkgs/main/noarch::tzdata-2022g-h04d1e81_0
urllib3            pkgs/main/osx-arm64::urllib3-1.26.14-py310hca03da5_0
w3lib              pkgs/main/noarch::w3lib-1.21.0-pyhd3eb1b0_0
wheel              pkgs/main/noarch::wheel-0.37.1-pyhd3eb1b0_0
xz                 pkgs/main/osx-arm64::xz-5.2.10-h80987f9_1
zlib               pkgs/main/osx-arm64::zlib-1.2.13-h5a0b063_0
zope                pkgs/main/osx-arm64::zope-1.0-py310hca03da5_1
zope.interface      pkgs/main/osx-arm64::zope.interface-5.4.0-py310h1a28f6b_0

Proceed ([y]/n)? y

Preparing transaction: done
Verifying transaction: done
Executing transaction: done
(exp3) Rahuls-Air:~ rahulgusain$ scrapy startproject spyder
New Scrappy project 'spyder', using template directory '/opt/homebrew/Caskroom/miniconda/base/envs/exp3/lib/python3.10/site-packages/scrappy/templates/project', created in:
/Users/rahulgusain/spyder

You can start your first spider with:
  cd spyder
  scrapy genspider example example.com
(exp3) Rahuls-Air:~ rahulgusain$

```

Step 2: - Now we will create our first spider. To do that we need to create a python file named `quotes_spider.py` in `spyder/spyder/spiders` directory.

Command: - `cd spyder/spiders`

```

requests          pkgs/main/osx-arm64::requests-2.28.1-py310hca03da5_0
requests-file     pkgs/main/noarch::requests-file-1.5.1-pyhd3eb1b0_0
scrapy            pkgs/main/osx-arm64::scrapy-2.6.2-py310hca03da5_0
service_identity pkgs/main/noarch::service_identity-18.1.0-pyhd3eb1b0_1
setuptools        pkgs/main/osx-arm64::setuptools-65.6.3-py310hca03da5_0
six               pkgs/main/noarch::six-1.16.0-pyhd3eb1b0_1
sqlite            pkgs/main/osx-arm64::sqlite-3.40.1-h7a7dc30_0
tk                pkgs/main/osx-arm64::tk-8.6.12-hb8d0fd4_0
tldextract        pkgs/main/noarch::tldextract-3.2.0-pyhd3eb1b0_0
twisted           pkgs/main/osx-arm64::twisted-22.2.0-py310h1a28f6b_1
typing_extensions pkgs/main/osx-arm64::typing_extensions-4.4.0-py310hca03da5_0
tzdata            pkgs/main/noarch::tzdata-2022g-h04d1e81_0
urllib3           pkgs/main/osx-arm64::urllib3-1.26.14-py310hca03da5_0
w3lib             pkgs/main/noarch::w3lib-1.21.0-pyhd3eb1b0_0
wheel             pkgs/main/noarch::wheel-0.37.1-pyhd3eb1b0_0
xz                pkgs/main/osx-arm64::xz-5.2.10-h80987f9_1
zlib              pkgs/main/osx-arm64::zlib-1.2.13-h5a0b063_0
zope              pkgs/main/osx-arm64::zope-1.0-py310hca03da5_1
zope.interface    pkgs/main/osx-arm64::zope.interface-5.4.0-py310h1a28f6b_0

Proceed ([y]/n)? y

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

You can start your first spider with:
cd spyder
scrapy genspider example example.com
(exp3) Rahul-Air:~ rahulgusain$

```

```

(exp3) Rahul-Air:~ rahulgusain$ cd spyder
(exp3) Rahul-Air:spyder rahulgusain$ cd spyder
(exp3) Rahul-Air:spyder rahulgusain$ cd spiders
(exp3) Rahul-Air:spiders rahulgusain$ ls
__init__.py
(exp3) Rahul-Air:spiders rahulgusain$

```

Command: - touch quotes_spider.py

```

Two_Sum.py  sqrt.py 1  quotes.py 1 X
Users > rahulgusain > spyder > spyder > spiders > quotes.py > QuotesSpider > parse
9
10 def start_requests(self):
11     urls = [
12         'https://quotes.toscrape.com/page/1/',
13         'https://quotes.toscrape.com/page/2/'
14     ]
15     for url in urls:
16         yield scrapy.Request(url=url, callback=self.parse)
17
18 def parse(self, response):
19     page = response.url.split("/")[-2]
20     filename = f'quotes-{page}.html'
21     Path(filename).write_bytes(response.body)
22     self.log(f'Saved file {filename}')

```

PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL Filter (e.g. text, **/*.ts, !**/node_modules/**)

- sqrt.py 1
 - Expected indented block Pylance [Ln 3, Col 9]
- quotes.py ~/spyder/spyder/spiders 1
 - Import "scrapy" could not be resolved Pylance(reportMissingImports) [Ln 3, Col 8]

OUTLINE TIMELINE

Ln 21, Col 43 Spaces: 4 UTF-8 LF Python 3.11.0 64-bit Go Live

```

2023-02-10 11:33:16 [scrapy.core.engine] DEBUG: Crawled (404) <GET https://quotes.toscrape.com/robots.txt> (referer: None)
2023-02-10 11:33:16 [scrapy.core.engine] DEBUG: Crawled (200) <GET https://quotes.toscrape.com/page/1/> (referer: None)
2023-02-10 11:33:16 [quotes] DEBUG: Saved file quotes-1.html
2023-02-10 11:33:16 [scrapy.core.engine] DEBUG: Crawled (200) <GET https://quotes.toscrape.com/page/2/> (referer: None)
2023-02-10 11:33:17 [quotes] DEBUG: Saved file quotes-2.html
2023-02-10 11:33:17 [scrapy.core.engine] INFO: Closing spider (finished)
2023-02-10 11:33:17 [scrapy.statscollectors] INFO: Dumping Scrapy stats:
{
  'downloader/request_bytes': 693,
  'downloader/request_count': 3,
  'downloader/request_method_count/GET': 3,
  'downloader/response_bytes': 5557,
  'downloader/response_count': 3,
  'downloader/response_status_count/200': 2,
  'downloader/response_status_count/404': 1,
  'elapsed_time_seconds': 3.74965,
  'finish_reason': 'finished',
  'finish_time': datetime.datetime(2023, 2, 10, 6, 3, 17, 75874),
  'httpcompression/response_bytes': 25019,
  'httpcompression/response_count': 3,
  'log_count/DEBUG': 6,
  'log_count/INFO': 10,
  'memusage/max': 69500928,
  'memusage/startup': 69500928,
  'response_received_count': 3,
  'robotstxt/request_count': 1,
  'robotstxt/response_count': 1,
  'robotstxt/response_status_count/404': 1,
  'scheduler/dequeued': 2,
  'scheduler/dequeued/memory': 2,
  'scheduler/enqueued': 2,
  'scheduler/enqueued/memory': 2,
  'start_time': datetime.datetime(2023, 2, 10, 6, 3, 13, 326224)}
2023-02-10 11:33:17 [scrapy.core.engine] INFO: Spider closed (finished)
(exp3) Rahuls-Air:spiders rahulgusain$

```

Step 3: - Now we will store the scrapped data into json file and csv file. For that we need to execute the below command.

Command: - scrapy crawl quotes -O quotes.json

Command: - scrapy crawl quotes -O quotes.csv

```

(exp3) Rahuls-Air:spiders rahulgusain$ touch quotes.py
(exp3) Rahuls-Air:spiders rahulgusain$ scrapy crawl quotes
2023-02-10 11:33:10 [scrapy.utils.log] INFO: Scrapy 2.6.2 started (bot: spyder)
2023-02-10 11:33:10 [scrapy.utils.log] INFO: Versions: lxml 4.9.1.0, libxml2 2.9.14, cssselect 1.1.0, parsel 1.6.0, w3lib 1.21.0, Twisted 22.2.0, Python 3.10.9 (main, Jan 11 2023, 09:18:18) [Clang 14.0.6 ], pyOpenSSL 22.0.0 (OpenSSL 1.1.1s 1 Nov 2022), cryptography 38.0.4, Platform macOS-13.0.1-arm64-arm-64bit
2023-02-10 11:33:10 [scrapy.crawler] INFO: Overridden settings:
{'BOT_NAME': 'spyder',
 'NEWSPIDER_MODULE': 'spyder.spiders',
 'ROBOTSTXT_OBEY': True,
 'SPIDER_MODULES': ['spyder.spiders']}
2023-02-10 11:33:10 [scrapy.utils.log] DEBUG: Using reactor: twisted.internet.selectreactor.SelectReactor
2023-02-10 11:33:10 [scrapy.extensions.telnet] INFO: Telnet Password: 33d82eca6667198a
2023-02-10 11:33:12 [scrapy.middleware] INFO: Enabled extensions:
['scrapy.extensions.corestats.CoreStats',
 'scrapy.extensions.telnet.TelnetConsole',
 'scrapy.extensions.memusage.MemoryUsage',
 'scrapy.extensions.logstats.LogStats']
2023-02-10 11:33:12 [scrapy.middleware] INFO: Enabled downloader middlewares:
['scrapy.downloadermiddlewares.robotstxt.RobotsTxtMiddleware',
 'scrapy.downloadermiddlewares.httpauth.HttpAuthMiddleware',
 'scrapy.downloadermiddlewares.downloadtimeout.DownloadTimeoutMiddleware',
 'scrapy.downloadermiddlewares.defaultheaders.DefaultHeadersMiddleware',
 'scrapy.downloadermiddlewares.useragent.UserAgentMiddleware',
 'scrapy.downloadermiddlewares.retry.RetryMiddleware',
 'scrapy.downloadermiddlewares.redirect.MetaRefreshMiddleware',
 'scrapy.downloadermiddlewares.httpcompression.HttpCompressionMiddleware',
 'scrapy.downloadermiddlewares.redirect.RedirectMiddleware',
 'scrapy.downloadermiddlewares.cookies.CookiesMiddleware',
 'scrapy.downloadermiddlewares.httpproxy.HttpProxyMiddleware',
 'scrapy.downloadermiddlewares.stats.DownloaderStats']
2023-02-10 11:33:12 [scrapy.middleware] INFO: Enabled spider middlewares:
['scrapy.spidermiddlewares.httperror.HttpErrorMiddleware',
 'scrapy.spidermiddlewares.offsite.OffsiteMiddleware',
 'scrapy.spidermiddlewares.referer.RefererMiddleware',

```

```
[ 'scrapy.spidermiddlewares.httperror.HttpErrorMiddleware',
  'scrapy.spidermiddlewares.offsite.OffsiteMiddleware',
  'scrapy.spidermiddlewares.referer.RefererMiddleware',
  'scrapy.spidermiddlewares.urllength.UrlLengthMiddleware',
  'scrapy.spidermiddlewares.depth.DepthMiddleware']
2023-02-10 11:33:53 [scrapy.middleware] INFO: Enabled item pipelines:
[]
2023-02-10 11:33:53 [scrapy.extensions.telnet] INFO: Telnet console listening on 127.0.0.1:6023
2023-02-10 11:33:53 [scrapy.core.engine] INFO: Spider opened
2023-02-10 11:33:56 [scrapy.core.engine] DEBUG: Crawled (404) <GET https://quotes.toscrape.com/robots.txt> (referer: None)
2023-02-10 11:33:56 [scrapy.core.engine] DEBUG: Crawled (200) <GET https://quotes.toscrape.com/page/1/> (referer: None)
[s] Available Scrapy objects:
[s] scrapy scrapy module (contains scrapy.Request, scrapy.Selector, etc)
[s] crawler <scrapy.crawler.Crawler object at 0x10618f550>
[s] item {}
[s] request <GET https://quotes.toscrape.com/page/1/>
[s] response <200 https://quotes.toscrape.com/page/1/>
[s] settings <scrapy.settings.Settings object at 0x10618f1c0>
[s] spider <DefaultSpider 'default' at 0x108329930>
[s] Useful shortcuts:
[s] fetch(url[, redirect=True]) Fetch URL and update local objects (by default, redirects are followed)
[s] fetch(req) Fetch a scrapy.Request and update local objects
[s] shell() Shell help (print this help)
[s] view(response) View response in a browser
>>> response.css('.text::text').getall()
["The world as we have created it is a process of our thinking. It cannot be changed without changing our thinking.",
 "It is our choices, Harry, that show what we truly are, far more than our abilities.", "There are only two ways to
live your life. One is as though nothing is a miracle. The other is as though everything is a miracle.", "The person,
be it gentleman or lady, who has not pleasure in a good novel, must be intolerably stupid.", "Imperfection is beauty
, madness is genius and it's better to be absolutely ridiculous than absolutely boring.", "Try not to become a man of
success. Rather become a man of value.", "It is better to be hated for what you are than to be loved for what you ar
e not.", "I have not failed. I've just found 10,000 ways that won't work.", "A woman is like a tea bag; you never k
now how strong it is until it's in hot water.", "A day without sunshine is like, you know, night." ]
>>>

{'item_scraped_count': 20,
 'log_count/DEBUG': 24,
 'log_count/INFO': 11,
 'memusage/max': 69582848,
 'memusage/startup': 69582848,
 'response_received_count': 3,
 'robotstxt/request_count': 1,
 'robotstxt/response_count': 1,
 'robotstxt/response_status_count/404': 1,
 'scheduler/dequeued': 2,
 'scheduler/dequeued/memory': 2,
 'scheduler/enqueued': 2,
 'scheduler/enqueued/memory': 2,
 'start_time': datetime.datetime(2023, 2, 10, 6, 11, 57, 165106)}
2023-02-10 11:42:01 [scrapy.core.engine] INFO: Spider closed (finished)
(exp3) Rahul-Air:spiders rahulgusain$ scrapy startproject images
New Scrapy project 'images', using template directory '/opt/homebrew/Caskroom/miniconda/base/envs/exp3/lib/python3.10/s
ite-packages/scrapy/templates/project', created in:
/Users/rahulgusain/spyder/spyder/spiders/images

You can start your first spider with:
cd images
scrapy genspider example example.com
(exp3) Rahul-Air:spiders rahulgusain$ cd
__init__.py __pycache__ images/ quotes-1.html quotes-2.html quotes.csv quotes.py
(exp3) Rahul-Air:spiders rahulgusain$ cd images
(exp3) Rahul-Air:images rahulgusain$ ls
images scrapy.cfg
(exp3) Rahul-Air:images rahulgusain$ cd images
(exp3) Rahul-Air:images rahulgusain$ ls
__init__.py items.py middlewares.py pipelines.py settings.py spiders
(exp3) Rahul-Air:images rahulgusain$ cd spiders
(exp3) Rahul-Air:spiders rahulgusain$ ls
__init__.py
(exp3) Rahul-Air:spiders rahulgusain$
```

