

# Project: Parallel Web Crawler

## 7.2 Getting Started

In terminal

```
$ cd starter/
```

```
$ cd webcrawler/
```

```
$ mvn test -Dtest=WebCrawlerTest
```

```
-DcrawlerImplementations=com.udacity.webcrawler.SequentialWebCrawler
```

```
[INFO] -----  
[INFO] T E S T S  
[INFO] -----  
[INFO] Running com.udacity.webcrawler.WebCrawlerTest  
[INFO] Tests run: 13, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 1.359 s - in com.udacity.webcrawler.WebCrawlerTest  
[INFO]  
[INFO] Results:  
[INFO]  
[INFO] Tests run: 13, Failures: 0, Errors: 0, Skipped: 0  
[INFO]  
[INFO] -----  
[INFO] BUILD SUCCESS  
[INFO] -----  
[INFO] Total time: 10.231 s  
[INFO] Finished at: 2024-10-07T23:57:43+05:30  
[INFO] -----
```

## 7.3 Crawler Configuration

In terminal

```
$ mvn test -Dtest=ConfigurationLoaderTest
```

```
[INFO] -----  
[INFO] T E S T S  
[INFO] -----  
[INFO] Running com.udacity.webcrawler.json.ConfigurationLoaderTest  
[INFO] Tests run: 2, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.775 s - in com.udacity.webcrawler.json.ConfigurationLoaderTest  
[INFO]  
[INFO] Results:  
[INFO]  
[INFO] Tests run: 2, Failures: 0, Errors: 0, Skipped: 0  
[INFO]  
[INFO] -----  
[INFO] BUILD SUCCESS  
[INFO] -----  
[INFO] Total time: 9.368 s  
[INFO] Finished at: 2024-10-08T00:14:57+05:30  
[INFO] -----
```

## 7.4 Crawler Output

In terminal

```
$ mvn test -Dtest=CrawlResultWriterTest
```

```
[INFO] -----  
[INFO] T E S T S  
[INFO] -----  
[INFO] Running com.udacity.webcrawler.json.CrawlResultWriterTest  
[INFO] Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.742 s - in com.udacity.webcrawler.json.CrawlResultWriterTest  
[INFO]  
[INFO] Results:  
[INFO]  
[INFO] Tests run: 1, Failures: 0, Errors: 0, Skipped: 0  
[INFO]  
[INFO] -----  
[INFO] BUILD SUCCESS  
[INFO] -----  
[INFO] Total time: 4.834 s  
[INFO] Finished at: 2024-10-08T00:16:40+05:30  
[INFO] -----
```

## 7.5 Legacy Crawler

In terminal

```
$ mvn package '-Dmaven.test.skip=true'
```

```
[INFO] -----  
[INFO] BUILD SUCCESS  
[INFO] -----  
[INFO] Total time: 6.945 s  
[INFO] Finished at: 2024-10-08T00:28:32+05:30  
[INFO] -----
```

```
$ java -classpath target/udacity-webcrawler-1.0.jar  
com.udacity.webcrawler.main.WebCrawlerMain  
src/main/java/com/udacity/webcrawler/main/config/sample_config_sequential.json
```

```
{"wordCounts":{"borrow":124,"library":35,"books":27,"read":16,"open":12},"urlsVisited":1}Run at Mon, 7 Oct 2024 19:00:43 GMT  
com.udacity.webcrawler.SequentialWebCrawler#crawl took 0m 8s 603ms  
com.udacity.webcrawler.parser.ParserImpl#parse took 0m 8s 743ms
```

## 7.6 Parallel Crawler

In terminal

```
$ mvn test -Dtest=WebCrawlerTest,ParallelWebCrawlerTest
```

```

[INFO] -----
[INFO] T E S T S
[INFO] -----
[INFO] Running com.udacity.webcrawler.ParallelWebCrawlerTest
[INFO] Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.853 s - in com.udacity.webcrawler.ParallelWebCrawlerTest
[INFO] Running com.udacity.webcrawler.WebCrawlerTest
[INFO] Tests run: 22, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.725 s - in com.udacity.webcrawler.WebCrawlerTest
[INFO]
[INFO] Results:
[INFO]
[INFO] Tests run: 23, Failures: 0, Errors: 0, Skipped: 0
[INFO]
[INFO] -----
[INFO] BUILD SUCCESS
[INFO] -----
[INFO] Total time: 5.655 s
[INFO] Finished at: 2024-10-08T00:35:50+05:30
[INFO] -----

```

## 7.7 Functional Counting

In terminal

```
$ mvn test -Dtest=WordCountsTest
```

```

[INFO] -----
[INFO] T E S T S
[INFO] -----
[INFO] Running com.udacity.webcrawler.WordCountsTest
[INFO] Tests run: 2, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.173 s - in com.udacity.webcrawler.WordCountsTest
[INFO]
[INFO] Results:
[INFO]
[INFO] Tests run: 2, Failures: 0, Errors: 0, Skipped: 0
[INFO]
[INFO] -----
[INFO] BUILD SUCCESS
[INFO] -----
[INFO] Total time: 6.856 s
[INFO] Finished at: 2024-10-08T00:37:18+05:30
[INFO] -----

```

## 7.8 Performance Profiler

In terminal

```
$ mvn test -Dtest=ProfilerImplTest
```

```

[INFO] -----
[INFO] T E S T S
[INFO] -----
[INFO] Running com.udacity.webcrawler.profiler.ProfilerImplTest
[INFO] Tests run: 7, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.261 s - in com.udacity.webcrawler.profiler.ProfilerImplTest
[INFO]
[INFO] Results:
[INFO]
[INFO] Tests run: 7, Failures: 0, Errors: 0, Skipped: 0
[INFO]
[INFO] -----
[INFO] BUILD SUCCESS
[INFO] -----
[INFO] Total time: 4.975 s
[INFO] Finished at: 2024-10-08T00:38:46+05:30
[INFO] -----

```

## \$ mvn test

```

[INFO] -----
[INFO] T E S T S
[INFO] -----
[INFO] Running com.udacity.webcrawler.SequentialWebCrawlerTest
[INFO] Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.881 s - in com.udacity.webcrawler.SequentialWebCrawlerTest
[INFO] Running com.udacity.webcrawler.ParallelWebCrawlerTest
[INFO] Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.036 s - in com.udacity.webcrawler.ParallelWebCrawlerTest
[INFO] Running com.udacity.webcrawler.parser.PageParserImplTest
[INFO] Tests run: 2, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.258 s - in com.udacity.webcrawler.parser.PageParserImplTest
[INFO] Running com.udacity.webcrawler.json.ConfigurationLoaderTest
[INFO] Tests run: 2, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.734 s - in com.udacity.webcrawler.json.ConfigurationLoaderTest
[INFO] Running com.udacity.webcrawler.json.CrawlResultWriterTest
[INFO] Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.045 s - in com.udacity.webcrawler.json.CrawlResultWriterTest
[INFO] Running com.udacity.webcrawler.WebCrawlerTest
[INFO] Tests run: 22, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.898 s - in com.udacity.webcrawler.WebCrawlerTest
[INFO] Running com.udacity.webcrawler.profiler.ProfilerImplTest
[INFO] Tests run: 7, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.184 s - in com.udacity.webcrawler.profiler.ProfilerImplTest
[INFO] Running com.udacity.webcrawler.WordCountsTest
[INFO] Tests run: 2, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.049 s - in com.udacity.webcrawler.WordCountsTest
[INFO]
[INFO] Results:
[INFO]
[INFO] Tests run: 38, Failures: 0, Errors: 0, Skipped: 0
[INFO]
[INFO] -----
[INFO] BUILD SUCCESS
[INFO] -----
[INFO] Total time: 7.806 s
[INFO] Finished at: 2024-10-08T00:40:07+05:30
[INFO] -----

```

## 7.9 Let's Run It!

In terminal

\$ mvn package

```
[INFO] -----  
[INFO] BUILD SUCCESS  
[INFO] -----  
[INFO] Total time: 12.452 s  
[INFO] Finished at: 2024-10-08T00:44:38+05:30  
[INFO] -----
```

```
$ java -classpath target/udacity-webcrawler-1.0.jar  
com.udacity.webcrawler.main.WebCrawlerMain  
src/main/java/com/udacity/webcrawler/main/config/sample_config.json
```

```
sample_config.json ×  
1  {  
2    "startPages": ["https://www.google.org/"],  
3    "ignoredUrls": [  
4      "https://www.google.org/our-approach.*"  
5    ],  
6    "ignoredWords": ["^.{1,3}$"],  
7    "parallelism": 4,  
8    "maxDepth": 2,  
9    "timeoutSeconds": 5,  
10   "popularWordCount": 5,  
11   "profileOutputPath": "profileData.txt"  
12 }
```

```
{"wordCounts":{"google":701,"your":559,"with":293,"services":246,"more":235},"urlsVisited":19}
```