

Hrishikesh Sanjay Pawar – SEC01 (NUID 002707307)

Big Data System Engineering with Scala

Spring 2023

Assignment No. 6



-List of Tasks Implemented

Completed all the incomplete implementations to get the web crawler working

-Code

```
134  /**
135   * Method to get a list of URLs referenced by the given URL.
136   *
137   * @param url a URL.
138   * @return a Future of Seq[URL] which corresponds to the various A links in the HTML.
139   */
140  def wget(url: URL)(implicit ec: ExecutionContext): Future[Seq[URL]] = {
141    // Hint: write as a for-comprehension, using the method createURL(Option[URL], String) to get the appropriate URL for relative links
142    // 16 points.
143
144    def getURLs(ns: Node): Seq[Try[URL]] = {
145      // (ns \\ "a").flatMap(_ \\ "@href").map(href => validateURL(createURL(Some(url), href.toString)))
146
147      for {
148        a <- ns \\ "a";
149        x <- a \\ "@href"
150      } yield validateURL(createURL(Some(url), x.toString))
151    }
```

```
153  def getLinks(g: String): Try[Seq[URL]] = {
154    val ny: Try[Node] = HTMLParser.parse(g) recoverWith { case f => Failure(new RuntimeException(s"parse problem with URL $url: $f")) }
155    for (n <- ny; uys = getURLs(n); us <- MonadOps.sequenceForgiveSubsequent(uys) { case _: WebCrawlerProtocolException => true; case _ => false })
156  }
157  // Hint: write as a for-comprehension, using getURLContent (above) and getLinks above. You will also need MonadOps.asFuture
158  // 9 points.
159  for {
160    sf <- getURLContent(url);
161    links <- MonadOps.asFuture(getLinks(sf))
162  } yield links // TO BE IMPLEMENTED
163 }
```

```
240  /**
241   * Method to extract an Option[X] from an Either[Throwable, X].
242   *
243   * Hint: this one is easy: just look for a method which turns an Either into a Option.
244   * 7 points.
245   *
246   * @param xe an Either[Throwable, X].
247   * @param X the underlying type.
248   * @return if xe is a Right(x) then Some(x) else None.
249   */
250  def asOption[X](xe: Either[Throwable, X]): Option[X] = xe.toOption
251 }
```

-Unit tests

Paste Screenshots of your unit tests. (Make sure to include the time on your screen)

Incase Screenshot is not clear, paste a screenshot of all unit tests separately and a screenshot of the entire screen with the time included.

CSYE7200 - MonadOpsSpec.scala [WebCrawler]

Hrishikesh-Pawar | assignment-web-crawler | src | test | scala | edu | neu | coe | csye7200 | asstwc | QueueOpsSpec.scala | Current File | Git: ✓ ✓ ✓ ✓ ✓

```
13 import ...
14 /**
15  * T000 implement me properly
16  */
17 class MonadOpsSpec extends flatspec.AnyFlatSpec with should.Matchers with Futures with ScalaFutures {
18
19   import ...
20
21   behavior of "LiftFuture"
22   it should "work" in {
23     val g: Future[Int] => Future[String] = liftFuture(_._toString)
24   }
```

Run: MonadOpsSpec

Tests passed: 31 of 31 tests - 2 sec 130 ms

Test Results

Test	Duration
MonadOpsSpec	2 sec 130 ms
LiftFuture	2 sec 130 ms
should work	124 ms
AsFuture	124 ms
SequenceForgivingWithLogging	5 ms
SequenceWithLogging	20 ms
sequenceLax	8 ms
SequenceForgiving	6 ms
LiftTry	7 ms
LiftTry	6 ms

Tests passed: 31 (moments ago)

3°C Clear

21:1 CRLF UTF-8 2 spaces Assignment6-Fall2022 01:57 21-03-2023

CSYE7200 - QueueOpsSpec.scala [WebCrawler]

Hrishikesh-Pawar | assignment-web-crawler | src | test | scala | edu | neu | coe | csye7200 | asstwc | QueueOpsSpec.scala | Current File | Git: ✓ ✓ ✓ ✓ ✓

```
1 package edu.neu.coe.csye7200.asstwc
2
3 import ...
4
5 class QueueOpsSpec extends AnyFlatSpec with should.Matchers {
6
7   behavior of "RichQueue"
8
9   it should "dequeueN(1) empty" in {
10     val queue = Queue()
11     val (x, z) = queue.dequeueN(1)
12   }
```

Run: QueueOpsSpec

Tests passed: 4 of 4 tests - 108 ms

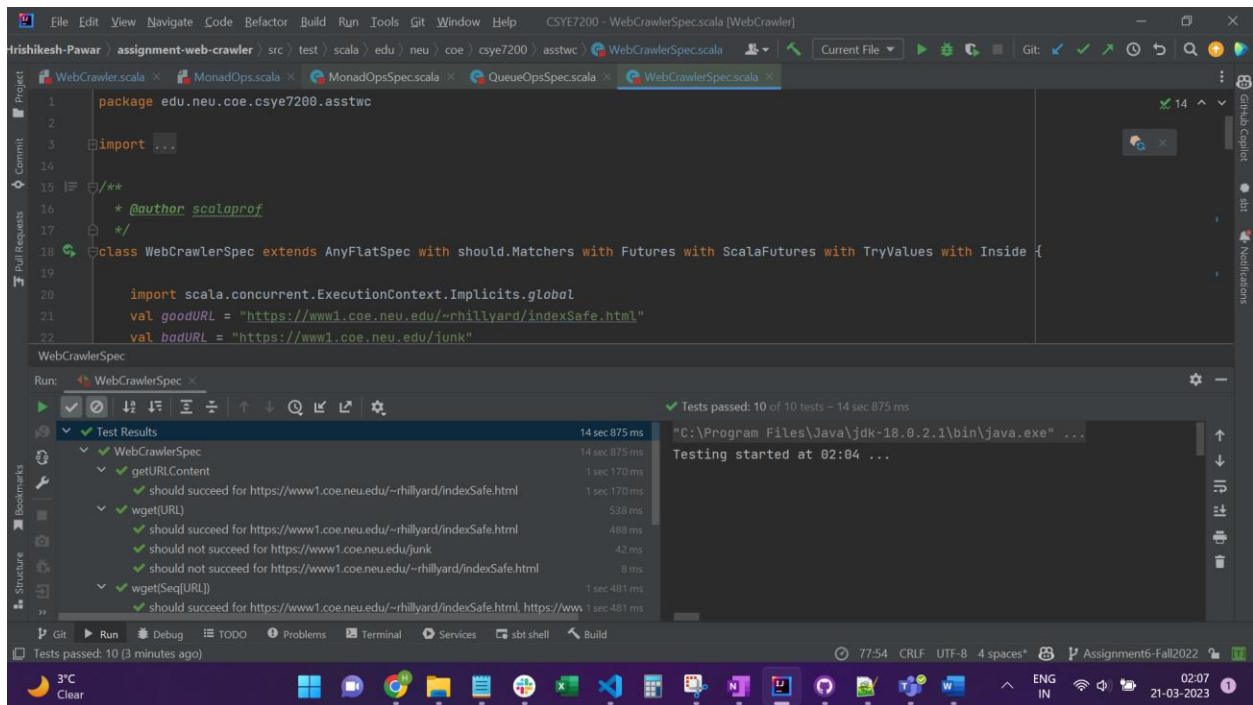
Test Results

Test	Duration
QueueOpsSpec	108 ms
RichQueue	108 ms
should dequeueN(1) empty	108 ms
should dequeueN(1)	85 ms
should dequeueN(2) A	1 ms
should dequeueN(2) B	16 ms
should dequeueN(2) B	6 ms

Tests passed: 4 (moments ago)

3°C Clear

8:7 CRLF UTF-8 4 spaces Assignment6-Fall2022 01:57 21-03-2023



- Result

Web crawler was successfully implemented.

- GitHub Repo

<https://github.com/hrishikesh-pawar/CSYE7200-Hrishikesh-Pawar/tree/Assignment6-Fall2022>