**Jawad Timzit**

**CSSE 436: Cloud Computing**

**Program 2: http-Program1**

This program takes in two arguments, an http and a NumofHops as a command line argument and will return http websites. My program will download html from the starting URL as the first argument and it will parse it finding the first <a href> reference to other absolute URL's. My program also will visit only a page 1 time if encounters a page that is been visited already it will skip and find the next reference. The operation will be repeated until there is no more links then it will print the html in text of the last page and the URL of each page that was visited. The program will be able to do the operation NumHops times.

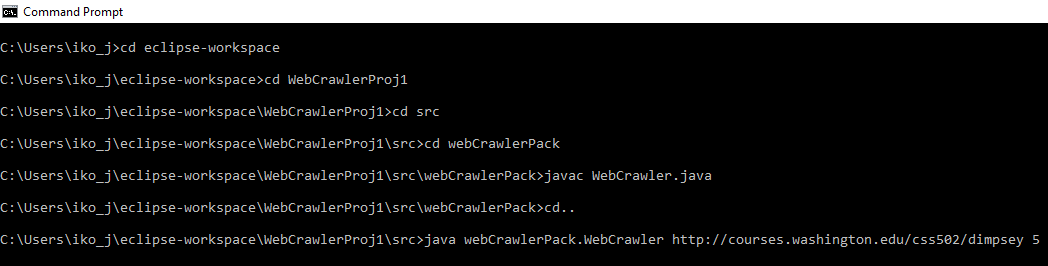
**\*\*\* COMPILATION INSTRUCTIONS \*\*\***

All necessary .java and .jar files for compiling the program are included in the project submission:

- WebCrawler.java

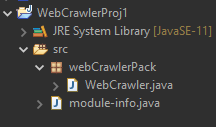
- WebCrawler.jar

In order to compile .java program, follow these same exact steps to compile the program using cmd in windows:



It’s necessary that you create a project – package - and a .java class to run those commands: (I used Jdk 11 for Java-you might won’t be able to run the program if you have an old version)

* You can use same name so it will be very easy to follow.
* Last line executes the program by taking in an URL and Num of Hops
* The console should then display the requested and display html of last page.



.

**Another way using Jar file:**

After you access the directory where you will save the .jar file

Run this command using cmd on windows: java -jar WebCrawler.jar <http://(some> URL) NumofHops. I have attached the manifesto with main expression on it.

* This one could be tricky because the manifesto file sometimes does not have manifesto with Main and does not run the program. It will return an error. So, make sure main is there on manifesto.mf otherwise you will have to edit it and add it.
* Expression to add to Manifesto second like if it was removed.

Main Class: webcrawlerpack.WebCrawler