SWINBURNE UNIVERSITY OF TECHNOLOGY

TNE20003 – Internet and Cybersecurity for Engineering Applications

Portfolio Task – Lab 7 Distinction Task

Aims:

- To develop a network client program to connect to a remote system and download data/information. Then to parse that information to extract data. Finally act on the downloaded data to fetch further information

Preparation:

- View "Internet Enabled Programming"

Due Date:

- Your Task will be assessed via an online quiz. You must score the required minimum to pass the test. You will be allowed a number of attempts to pass the test at the grade level you attempt. You are encouraged to complete the test at the end of the lab but if you do not, you must complete it before your next lab class. You must demonstrate the output to your lab demonstrator or if you don't get time upload it to Canvas along with your code.

Task 1

Take your completed CREDIT task and ensure that it is functioning as expected. We will now build onto that code by extracting further useful data.

You will be extending your program from the Credit task to parse the HTML data and search for any embedded HTML image tags. For each image embedded in the HTML, you need to download that image and save it to a disk file.

Task 2

You will need to extend your Python program in a number of ways:

- Ideally convert the code that downloads a URL to HTML into a function so that it can be called a second time to download images
- Parse the returned HTML code to extract any <img> tags
- For each image tag
    - Extract the path to download the image
    - From each path, extract the actual filename of the image
    - Construct a new URL to download the image and call your fetch function to download
    - Open a file in the program directory with the image name
    - Save the downloaded contents into that file

Hints:

As images are binary, you may want to ensure your fetch function works in binary and only convert to strings if required for the HTML and header parsing
Check your downloaded images by opening the file and seeing if it is a valid image