

**Name:** Sahand Sabour**Student ID:** 1614650**Project description**

This project implements the following parts of the project requirements:

- 1) Connect to a supplied URL and request the webpage and create the corresponding folders accordingly.
- 2) Download all images in the page. The program detects each image's path and the server that the image is stored on and then proceeds to act correspondingly. The mentioned image would later be saved in the folder for the page it was found on.
- 3) For all the hyperlinks in the page, the program would do step 1 and 2.
- 4) Supports threading and parallel download for images within a webpage.
- 5) Supports SSL connections.

**Guide for program use**

It should be noted that the entered URL must start with either `http://` or `https://` and it can only include spaces at the beginning and the end of it. As for the depth, user must not enter negative values. The program would crash otherwise.

**Known bugs**

- 1) The threading feature causes the printed statements to collide, which distorts the user interface. Moreover, a large number of images on a webpage interrupts the process.
- 2) Invalid input (for both depth and URL) causes the program to crash.
- 3) The proper functionality of the program depends on how a website implements its directories. For a hyperlink, it was assumed that if the link starts with a slash, it is opened from the main (home) directory and else, it exists in the current directory. Any website that uses a different implementation causes the program to fail.
- 4) Any website that doesn't use utf-8 encoding in their response crashes the program.
- 5) Accessing `https://` websites on systems in the XJTLU lab may crash the program.
- 6) If a webpage doesn't exist or is not found, its folder would still be created.
- 7) Some websites may disconnect the connection to the program because of time out and/or large number of requests.

**Resources**

Since none of the parts of the code that were inspired from other resources for this project exceed 10 lines of code, there are no resources (except for CSE205 lecture slides). However, the regular expressions used in this program were inspired by the [stackoverflow.com](https://stackoverflow.com) forums.