**Team Members: Yikai Liu, Calvin Jenks, Jay Patel, Khanh Le, Khang Truong**

# Multi-threaded Web Crawler

## Project Overview:

This project aims to build a powerful multi-threaded web crawler that can thoroughly explore all the links within a selected website. It's geared towards creating a detailed output, going up to a depth of 2. The tool is handy for making site maps or analyzing websites. To achieve this, we will use Jsoup, java.net, and java.util libraries to methodically go through each webpage, gather key details like titles and URLs and organize the findings into a structured output.

## Details:

The project uses Jsoup, java.net, and java.util libraries for the development of a web crawler. These libraries facilitate the systematic exploration of a website, extraction of pertinent information, and compilation of a structured output.

## Project Purpose:

The tool's main purpose is to enable the creation of sitemaps and streamline website analysis. By systematically analyzing hyperlinks within a specified depth, it extracts valuable information like titles and URLs. The structured output it generates serves as a concise list of all associated hyperlinks with the primary website.

## Input:

The user specifies the starting point for the crawl (e.g., www.gsu.edu, www.reddit.com, www.wikipedia.com, etc.), serving as the initial URL for the web crawler.

Change the desired hyperlink in line 18.

## Output:

The program prints out the discovered hyperlinks and their titles within the specified maximum depth. This output is displayed on the console, presenting the URL and title of each valid hyperlink. It provides an organized format that can be utilized for various applications, such as creating tree structures or generating sitemaps.

## Applications:

The resulting output provides a list of all links associated with the primary website, serving as a valuable resource for various purposes. These could include organizing hierarchical structures or creating sitemaps, enhancing the understanding and evaluation of the website's layout and content.

To run the code, download the JetBrains Toolbox. Then choose IntelliJ IDEA Community Edition. Make a folder that contains your code locally (or from GitHub).

Then, download the Jsoup package named “jsoup-1.17.2.jar core library”. Once downloaded, in IntelliJ, right-click on the folder name, then select: Open Module Settings → Libraries → “plus” symbol → java. Choose the Jsoup package you just downloaded above. Finally, you can run the code.