Web Clawer 3S1G

Project Overview:
The main goal of this project is to create an advanced multi-threaded web crawler capable of systematically examining all hyperlinks within a chosen website. The focus is on generating a comprehensive output with a maximum depth of 2. This tool is designed for constructing sitemaps or aiding in website analysis. The implementation involves utilizing Jsoup, java.net, and java.util libraries to systematically traverse each website, extract relevant information such as titles and URLs, and compile a structured output.

Details:
The project utilizes Jsoup, java.net, and java.util libraries to create a web crawler. These libraries enable the systematic exploration of a website, extraction of key information, and the compilation of a structured output.

Project Purpose:
The primary purpose of the tool is to provide a means for constructing sitemaps and facilitating website analysis. The tool extracts valuable information such as titles and URLs by systematically analyzing hyperlinks within a specified depth. This structured output serves as a concise list of all associated hyperlinks with the primary website.

Input:
The user specifies the starting point for the crawl; for example, the school's website. This serves as the initial URL for the web crawler.

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

Applications:
The resulting output provides a brief list of all hyperlinks associated with the primary website, making it a valuable resource for additional applications. These applications may include the creation of tree structures or the generation of sitemaps, enhancing the overall understanding and analysis of the website's structure and content.

To run the code one must make sure one has a Java Development Kit as well as Jsoup installed, In this project we used IntelliJ IDEA CE to compile our code. Once the code is placed into the compiler, the Jsoup has to be installed as well. To do that you go to https://jsoup.org › download to download it. Once downloaded you go the Intelli and right-click on the folder's name under 'project' which is named Webclawer. Once you click you go down and click on open module settings. Once inside you click Libraries and hit the plus button, select java. This will open up your computer downloads and files, you want to find the Jsoup download from earlier and select it. Once selected it's added to the code. Now the code and run.