**DESIGN DOCUMENT**

**Steps for designing a document on how to crawl a website:**

Step 1: First we need to use cases and constraints and later we gather requirements and scope the problem.

Step 2: Now we will create a high-level design. Outline a high-level design with all important components.

Step 3: Now we will design core components. We will look into the details for each core component.

Step 4: Scale the design.

**Saving the images from the website**:

We will Right-click the URL, and choose Open in a new tab. In the new tab, right-click the image and Save the image then give the image a name and choose the destination on your pc to save it to.

**Saving the text from the website:**

Click the “File” menu and click the “Save as” or “Save Page As” option. Select “Web Page, HTML only” from the Save as Type drop-down menu, type a name for the file, and click “Save.” The text will be extracted and saved as an HTML file with the original page-formatting options intact.

The components I need while crawling the website are

Seed detector, Crawler Manager, Crawler, Link Extractor, Hypertext Analyzer, and HTTP Protocol Module.