Progress Report

1) Which tasks have been completed?

We have finished the code to scrape the URLs of BBC and CNN and their contents. generated files with a day's news from both sites. Additionally, for each of these two news sources, we created two separate files to store the information. One stores the news headlines and URLs, and the other stores the headlines and article contents.

2) Which tasks are pending?

We haven't completed the code to get information from other websites.

3) Are you facing any challenges?

When getting the URL of each news, it is a challenge to filter out non-text news URLs, such as live streaming and video URLs.