ENCE360 Assignment

Josiah Craw
35046080

October 18, 2019

1 Algorithm Analysis

1.1 Describe the Algorithm

The algorithm from lines 228-264 initially moves through each URL within the input file. It then proceeds to get the URL from the file by replacing the end character with a null terminator. The URL and the number of threads is then passed into get_num_tasks. This function destermins the number of tasks required to complete the download given the number of threads. The byte size is then returned. The number of tasks is used to loop through and add that number of tasks to the todo task queue. This queue contains tasks which hold the URL and the min and max range for one task. This separates the file into parts using the byte ranges. Next, the context containing the todo and complete queues is passed into wait_task. This function runs through all of the tasks in the queue and processes them using http_get_content. The files are then merged and the file fragments are deleted.

This algorithm is similar to the worker assignment algorithm that has been used in class.

1.2 Improvements

This algorithm could be improved by deleting the file fragments as they are merged into the final file. Running through all the files twice is inefficient. Splitting the files within threads would not necessarily increase efficiency however, this could help making the system if an internet failure occurs. For this improvement to be effective on running the code would have to check if the downloads had already occurred then catch up to the most recently downloaded file. Another improvement could be merging and deleting the files in a thread while starting the next download in the background. Alternatively the file fragments could be merged as they are downloaded individually.