Juan Hernandez

Journal

My first contribution to this project was to work on the insult filter. Our team talked about how we would store the HTML text and decided on making it a string. The program I worked on first looks for a text file “insults.txt” and then load the insult list into a struct array containing the insult list. If the file does not exist, the program then aborts. If the integration is complete, the program then moves on to a search and replace function to replace any words in the text if they are in the insults list.

One of the problems that I had with the search and replace was differentiations with upper/lowercase letters. For instance, if the word in the HTML string was “fUcK” the filter would fail. To prevent this, before the comparison takes place, the current string gets sent to a temporary string and then gets converted to lowercase. The insults list is in lowercase as well so then I just compare the two and change them if necessary. I originally had it comparing individual characters but realized that would filter out “good” words such as “assurance.” The next step is to throw the code into a header file and combine it with the main program.

After testing the filter, I encountered another problem. The way the filter was originally set up was to stream the string of text into tokens and check each token. This worked for filtering out the words but led to formatting issues due to the whitespace being ignored. What needed to be done is to add characters to a comparison string one by one and look ahead to see if whitespace or a new line character is next to signify the end of a word. Then check to see if the word is an insult and replace it if it is and adds the rest of the whitespace and new lines to the string to be returned until we get a character that isn’t and add it to the comparison string and start the cycle all over again.

A new issue arose after implementing this. The first word gets added to the comparison string correctly but afterwards, the comparison string has a whitespace added to the beginning of it. At first I thought it had something to do with the string.clear() call but after outputting the string and string length I found out that it was working correctly. I could not understand why this was happening since the comparison strings were being cleared after running the insult checker method. What I ended up doing to fix the error was to clear the comparison strings if a white-space was found in the data array. I still do not know how a white-space was added since there was an if statement implemented to make sure that did not happen and the strings were cleared.