

## HTML2Latex Report

The problem I set out to solve was that of converting an html file into a latex file with the same formatting. For this problem we were told to use the lex language and were provided a small skeleton program and a makefile. We were only required to support a subset of html tags and only some of them, `<p>` and `<li>` were allowed to contain other formatting tags.

Because of previous jobs and research I have work at I have a pretty good understanding of using regex. Since the main work of this project was to design the correct RegExs my approach was very simple. For the tags that allowed nesting I parse them in two parts. The first strips of the beginning tag and prints any unformatted text while the second part is a subloop that processes all nested tags. Once the closing tag is found I print any remaining unformatted text and then exit this subloop. This is done for both `<li>` and `<p>` and they both use the same subloop routines, but they have different expressions to determine if the end of the tag has been reached. The formatting tags we had to support were all relatively simple expressions and all had the same structure.

I tested my solution by simply working my way through the tags as they appeared in the test.html file and visually comparing my results to the expected outcome. I also used a regex testing site online when I was trying an idea that I ultimately did not end up using.

I did not have any major issue when doing this assignment aside from accidentally putting quotations in the wrong place and not being able to figure out why my expressions were not matching.