Sarah Huang
COSC 461: Intro to Compilers
September 20, 2022

html2latex Project Report

1. The problem I attempted to solve is to take a HTML file and translate the HTML tags and appearance into LaTeX. Once you get a tex file, the program will convert it into a PDF file.

2. First, I looked at the open-content LaTeX book and various online resources to get ideas on converting HTML tags to LaTeX because I was unfamiliar with the topic. The header tags and format text tags (small/big, bold, italics, etc.) are each contained in their own "simplified" function. Both front and end tags are in one statement, and the program prints out all the text between the tags.

```
"<h1>"[^<]*"</h1>" {
    yytext[yyleng-5] = 0;
    printf( "\\section{%s}", yytext+4 );
}
```

Other HTML tags like paragraphs, unordered/ordered lists, comments, and pre-formatted paragraphs require a specific environment. Rather than the front and end tags being together on one line, they are separated because it allows text format tags, list item tags, new lines, and whitespace to be included too.

```
"<pre>" {
    printf("\\begin{verbatim}");
    BEGIN PRE;
}
<PRE>"</pre>" {
    printf("\\end{verbatim}\n");
    BEGIN 0;
}
...
```

3. It was important to double-check the PDF output file and LaTeX commands once I finished converting the HTML tags to LaTeX. You can pipe the HTML file into the test executable and solution executable to print the LaTeX commands. Once the make test command is completed, I compared the PDF file to the original HTML file to verify the translation of HTML tags. Besides using the test.html file, I wrote my own basic HTML files to test each tag individually.

4. I was not familiar with LaTeX, so I had to some research to figure it out. Luckily, the header h1 and comment tags were already done, so it was a good starting point for me. It took me an embarrassing amount of time to realize that the program is printing LaTeX commands, and you need to add the right ones when certain HTML tags are read.