HTML2LaTeX

Synopsis

The HTML2LaTeX lab is an assignment where we have to use a combination of C and Lex to make an interpreter that accepts HTML as input, and converts it to LaTeX code. The solution to the problem given is fairly straight forward.

Approach

The solution to the problem given is pretty straight forward. I've been working with Regex for over a year now, and I have learned how to use it to its full extent. Unfortunately, this lab didn't allow me to use it to the extent I would have liked, but at least we got to use it. The obvious approach to the problem is to simply consider all possible cases of HTML tags given in the assignment, understand how HTML is written, and understand how LaTeX is written. After that, simply write the interpreter with Regex code to handle all of the capturing and replacing.
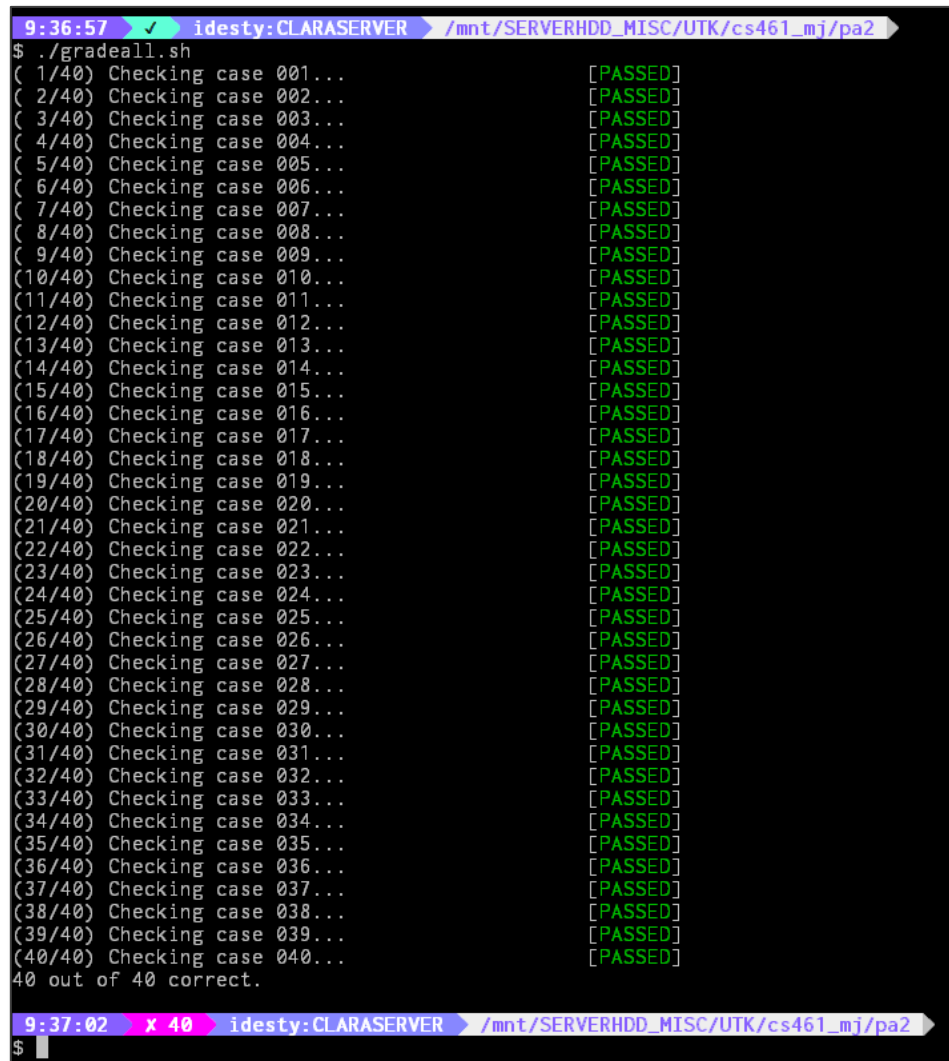
Debugging Solution

Inspired by the almighty Dr. Plank, I have decided to check my code via brute force by making test cases (which are actually included in this submitted tarball). I made 40 test cases (with a quarter of them being based on the test.html file given in the assignment tarball) and a few shell scripts. Executing "gradeall.sh" will run all 40 test cases on my executable and compare its output to cached output from the solution code. Executing "gradeall2.sh" will run all 40 test cases on both the solution code and my own code and compare their outputs. When I hit all 40 test cases, I knew that my code was good enough to submit. Please check page 2 to see a screenshot of the gradescript in action.

Issues

Coding in Lex for the first time was fairly (and disappointingly) trivial. There were not any problems at all with the assignment.

Gradescript

As mentioned above, here is a screenshot of the gradescript in action. It works on my personal server and on Tesla/Hydra. Feel free to try it yourself. More details on the shell scripts can be found in README.md.