COP701 Assignment 01

HTML to LaTeX Converter

Problem statement

This is your **first assignment** in the **COP 701** course.In this assignment your main objective is to convert a **HTML** document to an equivalent **LaTeX** document.
In pursuance of this objective, you will have to write a HTML to LaTeX parser from scratch.

The features(tags) of HTML which you all need to consider are:-

- **head**
- **body**
- **title**
- **a, href**
- **font: size**
- **center**
- **br**
- **p**
- **h1, h2, h3, h4**
- **ul, li, ol, ul, dl, dt, dd**
- **div**
- **u, b, i, em, tt, strong, small,**
- **sub, sup**
- **img: src, width, height, figure, figcaption**
- **table, caption, th, tr, td**

Some more commands, mathematical symbols, math-mode operators(for extra credit)

Workflow and subtasks

The entire assignment can be divided into the following sub-tasks:

- Learn about HTML and LaTeX in brief.
- Write a lexer i.e to do a lexical analysis of your HTML code and generate a string of tokens. Programs that you can use: flex, jflex
- Do **not** use any available libraries to parse the html.
- Parse the sequence of tokens using parser such as yacc, CUP, ANTLR, bison (C++ or Java)
- Generate an AST(Abstract Syntax Tree) of your HTML code. link
- Map it to an equivalent AST of LaTeX.
- Generate the equivalent LaTeX code which can be compiled to a PDF using TexMaker

Links to important resources

- **Know about Flex tool and performing lexical analysis using that.**
- **Some more resource regarding Flex.**
- **Introduction to an Abstract Syntax Tree.**
- **Bison parser generator**
- **YACC parser generator**

Logistics

- You are free to code in **any** programming language.
- The **deadline** for this assignment is **1/09/2019 at 11:55 pm**.It is a hard deadline and will not be extended.

- This is an individual assignment (30 Marks)
- Any form of **plagiarism** will not be tolerated.
- Also, create a **run.sh** file where the first argument will be the name of the html file and the second argument will be the name of the output tex file. We will run the command *./run.sh input.html output.tex* during the demo.
- Submission will be made on Moodle. You need to submit all your code (parser, translator) and a pdf format report. Compress all these in a tar file and upload on Moodle.
- You will be graded on the output of your code, the coding style and your viva/presentation.
- Marks distribution: Coding style - 25% , Demo - 75%
- We will be testing on hidden test cases during the demos

Sample test case

- sample1.html, img1 , img2