Report for the ASSIGNMENT 1 of SOFTWARE LAB COL 701 Akshay Gupta(MCS192556)

September 1, 2019

1 SOFTWARE LAB:HTML to LaTeX parser

1. The making of the lex file and generating tokens
   - The lex file that is made for generation of tokens is named "next.l"
   - At first HTML file is taken as input by the lex file for the generation of tokens
   - Then inside the lex file every tag is taken as an input and the specific token is generated of every tag
   - The tokens are generated for the DATA and Comment the same way as tokens for the tags were generated
   - The lex file is capable enough to identify the tokens and the type of tokens
   - the concepts of states like we used to do in TOC are also used to uniquely identify the type of tokens
   - The various regEx are used for idetifying the type token
   - Flex tool is used to compile the lex file
   - Now as the tokens are generated, now it will be passed from lex to yacc
2. Now some more information about the yacc file
   - The name of the yacc file used here is "next.y"
   - Now we will be getting the tokens from the lex file and they will be of different types
   - Here we actually assign the tokens type that should be passed from the lex file
   - The tokens type can be of (char *) or it can be of a type of tree node

- After assigning the types of tokens we will now write the grammer for it
- The grammer written will be conflicts free and will be unambiguous which will be effective while parsing.
- According to the HTML as understood the grammer is written in the same way to produce the productions
- BISON tool is used to execute the yacc file

3. What is actually stored in the AST

   - The header file containing all the function declarations is named as "next.h"
   - Now the AST will be made using functions declared in the header file.
   - newNode() function will be used for making the node of an AST
   - The enum nodetyp is used to take care of the nodes to which type do they belong
   - the different fields that is used to make the node are:
     (a) The DATA field
     (b) The Nodetype
     (c) It's Children
   - Different children of the node are stored using the vector
   - addchildren function is used for the adding of children to the existing node
   -

4. How is the AST is translated

   - The AST is translated using the depth first order traversal
   - It is traversed in inorder postorder and preorder all in combination to find the best suited solution for making of the file
   - The map is used for mapping of the HTML file to the equivalent LaTeX file
   - There are actually three map vectors that are used here for the mapping of the two equivalent nodes
   - startNodes and endNodes these are the two map vectors used for maping of the nodes
   - GreekWords is another map used to map the greek letters to their equivalent LaTeX form
   - seeImageAttribs function is used to find the image attributes in the node of type eIMAGE

- Same are the seeLinkAttribs, fontsizeprint
- seeTable function is used for the fetching of the whole table
- leftchild function is actually the traversing of the leftmost child
- passchildrenSKIP1 function is the exactly the opposite of leftchild function, it traverses every child except the leftmost child

Which Programming language is used

Here the programming language I used is C/C++11.