Report

September 1, 2019

1 LEX and YACC

1.1 LEX

- The Lex file contains the regular expressions for the tags mentioned.

- Four of the tags have been also checked for attributes namely Anchor Tag, Font Tag, Table Tag and Image Tag

- Initially many regular expressions are defined that are used in all the other regular expressions

- Some of these are {punc} which defines some punctuations, {spac} which defines spaces and tabs and {word} which uses {spac}, {punc} and other special characters to form contents of a word token