Thomas Miller and Jacob Whited

CS-4100

Homework Project 1

3/6/2023

For the lex portion, the program disregards comments, whitespace, and include statements as these are not the user-written program. It recognizes numbers and outputs them to the scanner\_out.txt file. Keywords "if", "else", "while", "for", "int", "char", "float", "double", "return", "void", "string", "bool", and "long" are output as strings containing the keyword. Operators "==", "!=", "<", ">", "<=", ">=", "&&", "||", "+", "-", "\*", "/", "=", "++", "--", and "%" are output as strings containing the operators. Brackets like “(”, “)”, “{”, and “}” are also tokenized as strings containing the operators. Function and variable names are tokenized as the general string “name” which serves to generalize surface level changes often made to attempt to hide plagiarism. The pattern is recognized when alphabetical characters are followed by numeric characters. Any combination of alphanumeric characters that starts with a letter is valid and tokenized as a name. All of the tokens are output on a single line with spaces in between them to the scanner\_out.txt file.

2. A brief description in your own words of how you implemented the Winnowing algorithm

3. The results of your analysis, including identifying any submissions found by the algorithm that

you believe may be plagiarism