ICSI 311 Assignment 2 – Enhance the Lexer

This assignment is extremely important – (nearly) every assignment after this one uses this one!

If you have bugs or missing features in this, you will need to fix them before you can continue on to new assignments. This is very typical in software development outside of school.

You must submit .java files. Any other file type will be ignored. Especially ".class" files.

You must not zip or otherwise compress your assignment. Blackboard will allow you to submit multiple files.

You must submit buildable .java files for credit.

This assignment **must** have three source files (Basic.java, Token.java, Lexer.java).

In this assignment, you will enhance your lexer to accept more symbols, getting ready to implement BASIC. Specifically, you will need to add:

= (token type **must** be EQUALS)

<

<=

>

>=

<> (NOTEQUALS)

(LPAREN

) RPAREN

"anything except a close quote" (STRING)

In addition, the lexer **must** accept words. Words are letters (capital, lowercase, or mixed) and can be optionally followed by any **one** of the following: $%

The lexer must keep a data structure (HashMap is suggested) of "known" words and their token types. Unknown tokens are given the token type "IDENTIFIER". We will add to the known words throughout the semester, but use "PRINT" (with PRINT as the token) as a first example.

If the lexer encounters a word followed by :, it should NOT be an identifier, but a "LABEL".

Rubric	Poor	OK	Good	Great
Comments	None/Excessive (0)	"What" not "Why", few (5)	Some "what" comments or missing some (7)	Anything not obvious has reasoning (10)
Variable/Function naming	Single letters everywhere (0)	Lots of abbreviations (5)	Full words most of the time (8)	Full words, descriptive (10)
Handle Symbols	Handles < 3 symbols (0)	Handles most character symbols (10)		Handles all symbols (20)
Allows Identifiers	Does not allow identifiers(0)	Doesn't properly tokenize identifiers (10)		Properly tokenizes (20)
Properly handles Known key words	Does not tokenize (0)	tokenizes as "IDENTIFIER" (10)	Properly tokenizes (20)	Properly tokenizes and is easily extensible (30)
Accepts labels	Does not allow labels (0)	Doesn't properly tokenize labels (5)		Properly labels (10)