BIL395: Programming Languages
Programming Assignment 1: Simple Lexical Analyzer

Deadline: Feb 13 at 8:00pm

Lexical Analysis is the first phase of a compiler. It converts a source code written in a high level program language into a sequence of lexemes along with their tokens.The task in this assignment is to write a simplified lexical analyzer that extracts lexemes from an input source code written in Java and determines their tokens as specified below.

Precisely, the first column of the following table shows the set of lexemes you need to recognize while the right shows their tokens.

Lexeme	**Token**
for	FOR_STATEMENT
(	LPARANT
)	RPARANT
int	INT_TYPE
char	CHAR_TYPE
=	ASSIGNM
;	SEMICOLON
>	GREATER
<	LESS
>=	GRE_EQ
<=	LESS_EQ
{	LCURLYB
}	RCURLYB
return	RETURN_STMT
-	SUBT
/	DIV
*	MULT
+	ADD
identifier	An identifier consists of a single letter
integer constant	INT_LIT

Sample input file and the output of the lexical analyzer:

Input:

```
for (int i = 0; i < 10; i=i+1)
     {
          a = f;
          char c;
     }
return 0;
```

Output:

```
Next token is FOR_STATEMENT        Next lexeme is for
Next token is LPARANT              Next lexeme is (
Next token is INT_TYPE             Next lexeme is int
Next token is identifier           Next lexeme is i
Next token is ASSIGNM              Next lexeme is =
Next token is INT_LIT              Next lexeme is 0
Next token is SEMICOLON            Next lexeme is ;
Next token is identifier           Next lexeme is i
Next token is LESS                 Next lexeme is <
Next token is INT_LIT              Next lexeme is 10
Next token is SEMICOLON            Next lexeme is ;
Next token is identifier           Next lexeme is i
Next token is ASSIGNM              Next lexeme is =
Next token is identifier           Next lexeme is i
Next token is ADD                  Next lexeme is +
Next token is INT_LIT              Next lexeme is 1
Next token is RPARANT              Next lexeme is )
Next token is LCURLYB              Next lexeme is {
Next token is identifier           Next lexeme is a
Next token is ASSIGNM              Next lexeme is =
Next token is identifier           Next lexeme is f
Next token is SEMICOLON            Next lexeme is ;
Next token is CHAR_TYPE            Next lexeme is char
Next token is identifier           Next lexeme is c
Next token is SEMICOLON            Next lexeme is ;
Next token is RCURLYB              Next lexeme is }
Next token is RETURN_STMT          Next lexeme is return
Next token is INT_LIT              Next lexeme is 0
Next token is SEMICOLON            Next lexeme is ;
```

If you want to know whether a particular source code can be given as an input to your program, please try it in a Java compiler. If it is accepted by Java then it can be given as an input. In any case, make sure you use only lexemes given in the table.

Your program should also check for errors. However, your first goal is to make sure that if a valid source code is given to your program as an input then it is correctly analyzed by your program.

Error types that could exist:

- Unknown operator: This occurs when an operator other than the ones given in the table is scanned. For example, !, @, #, $, % are all unknown operators
- Unknown identifier: This occurs when an identifier consisting of more than a single char exists in the input. For example, ab, ab1, xyz, while, do are unknown identifiers

Sample Run:

Your code will take input and output files from the console as given in the following example where the source code, input and output files are named `hw1_firstname_lastname.java`, `input.txt`, and `output.txt`, respectively:

```
java hw1_firstname_lastname.java input.txt output1.txt
```

Submission:

- Programs will be written in Java.
- This assignment must be done individually (No groups).
- Use the submission link on uzak.etu.edu.tr to send your assignment.
- DON'T compress your submissions.
- Only *.java files will be accepted.
- Submissions sent in different formats (e.g., TXT, ZIP, RAR, etc) will lose 10 points.
- Assignments submitted after the due date will receive 25 point deduction for each day following the due date.
- We'll use an online tool to compute the similarity between all submissions.
- Please see the Ethical Rules section on the syllabus before starting to implement the homework.