**CPSC 323 – 01 Compilers and Languages**

**Programming Assignment 1**

This assignment can be done in groups of 1-3 members.

Maximum points: **100**

*Problem Statement:*

Write a program to build a lexical analyzer.The lexical analyzer should accept a .txt file as an input, consisting of code snippets and comments, and it gives output by tokenizing the input file.

For example,

*Input:*

if (a >= b)

{

return 0;

} //check if a is greater than or equal to b

*Output:*

Set of lexemes and tokens (<lexemes> = <tokens>)

“

**“If”** = keyword

**“(“** = separator

**“a”** = identifier

**“>” =** operator

**“=”** = operator

**“b”** = identifier

“)” = separator

**“{“** = separator

**“return”** = keyword

**“0”** = integer

**“;”** = separator

**“}”** = separator

“

* You can use regular expressions from external resources for identification of lexemes.
* Your code should remove unnecessary items like white spaces and comment lines.

Your Submissions:

1. The file which consists of your code along with proper comments. [50 points]
2. Screenshot of your correct output. [20 points]
3. A **PDF** report where you must explain each line of your code along with proper reasoning and references and the total time and space complexities. The report should **not** exceed more than 2 pages. [30 points]

**You are free to choose any programming language between C++, Java, or Python.**