CS F363 Compiler Construction
Assignment-1 (Question-1)

Due date: 16 Feb 2023 11:59 PM **Marks : 8**

Given a regular expression r and an input string w, the LEX tool finds the longest prefix w_1 (say w_1 as a valid token) of the input w that can be generated from the regular expression r and repeats the process with the remaining part of the input w. If a character is not part of any valid token, it echoes. Therefore, each character of w is either part of a valid token or echoed. Your task is to write a C / C++ code that takes a regular expression r and a string w as an input and output the sequences of valid tokens, as per LEX tool, and echo if a character is not part of any token.

Note that here the tokens are actual lexemes.
Input format: A text file, *input.txt* contains two lines; the first line contains the regular expression and the second line contain the input string.
Please note that each sub-regular expression is parenthesized. See the examples given in the below.

Output format: Each valid token is preceded by a $ and each echo character is preceded by @.
A string s is a sequence of valid tokens and echo characters in the same order as they appear in the input string w and s ends with #.
Generate a file *output.txt* (do not use other names) that contains s.

Examples:

1. $r = ((((a)(a))^+)(b))$ and $w = aaaabaaabbaab$, then $s = \$aaaab@a\$aab@b\$aab\#$
2. $r = ((((a)(a))^*)(b))$ and $w = aaaabaaabbaab$, then $s = \$aaaab@a\$aab\$b\$aab\#$
3. $r = (((b)(((a)(a))^*))(b))$ and $w = baaaabbbbab$, then $s = \$baaaab\$bb@b@a@b\#$

Submission guidelines:

1. Submit a single C / C++ file and name it with your BITS ID.
2. DO NOT SUBMIT A FOLDER.
3. Strictly follow the input and output formats.
4. **Late submission:** Each 1 hr delay fetch 1% penalty and late submission will not be accepted after 48 hours from the due date.