CS F363 Compiler Construction
Assignment-1 (Question-2)

Due date: 28 Feb 2023 11:59 PM **Marks : 7**

Given n regular expressions $r_1, r_2, \ldots, r_n$ and a 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 by a regular expression (RE) in $\{r_1, r_2, \ldots, r_n\}$. If the longest prefix w_1 can be generated by more than one regular expression, consider that it is generated by the lower indexed regular expression among the REs that can generate w_1 and ouput the index of the regular expression preceded by a \$, i.e., if w_1 can be gerated by $r_3, r_5,$ and r_8, then consider that w_1 is generated by r_3 and output \$3. Now, repeats the process with the remaining part of the input w. If a character is not part of any valid token, it echoes so you output the charecter preceded by a @. 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 n regular expressions $r_1, r_2, \ldots, r_n$ and a string w 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.
Further note the following:

1. Assume that $\Sigma = \{a, b\}$ is the alphabet set i.e., any regular expression r (as given as input) will generate a language $L(r) \subseteq \{a, b\}^*$.
2. Operations on the regular expressions: concatenation, union, closures (both $*$ and $+$).

Input format: A text file, *input.txt* contains $n + 2$ lines; the fist line contain the value of n, the lines from $2, 3, \ldots, n + 1$ contains n regular expressions $r_1, r_2, \ldots, r_n$ (one per line) and the $(n + 2)$-th line contains the input string w.

Please note that each sub-regular expression is parenthesized as before.
Output format:
Generate a file, *output.txt* (do not use other names), that contains the output for the given instance.
Example :
content of input.txt:
2
$(((a)*)(b))$
$((b)(a))$
abbaababaaabbaa

content of output.txt: \$1\$2\$1\$1\$1\$2@*a*#

Few more test cases will be uploaded by 23 Feb 1 PM.

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.

You are not allowed to use <regex> library or any similar libraries