

# Tutorial on Lexical Analysis

February 17, 2022

1. Write regular expressions for the following languages :

- B • All strings of lower case letters that contain the five vowels in order.
- B • All strings of lowercase letter in which the letters are in ascending lexicographic order.
- B • Comments, consisting of a string surrounded by /\* and \*/ without an intervening \*/, unless it is inside double-quotes(")

Lec • All strings of digits with no repeated digits

Lec • All strings of digits with at most one repeated digits

2. A lexical analyzer uses the following patterns to recognize three tokens  $T_1$ ,  $T_2$  and  $T_3$  over the alphabet  $\{a, b, c\}$ .

$T_1 : a?(b|c)^*a$

$T_2 : b?(a|c)^*b$

$T_3 : c?(b|a)^*c$

Note also that the analyzer outputs the token that matches the longest possible prefix. If the string bbaacabc is processed by the analyzer, which one of the following is the sequence of tokens it outputs ?

- $T_1 T_2 T_3$
- $T_1 T_1 T_3$
- $T_3 T_3$

3. Unsigned numbers (integer or floating point) are strings such as, 5280, 0.1234, 6.336E4, or 1.89E-4. Write a regular expression which is a precise specification for this set of strings.