Tutorial 1 (Lexical Analysis)

1. It is often the case that two different regular expressions define the same set. Thus a^+ and aa^* both define the set of all strings containing only a (excluding λ). Consider the six regular expressions listed below. Which define the same set as a^+? For those that don't, explain why.

   a) $(a^*)^+$ b) $(a^+)^+$
   c) $(aa^*)^+$ d) $(aa^+)^+$
   e) $(a \mid a^*)^+$ f) $(a \mid a^+)^+$

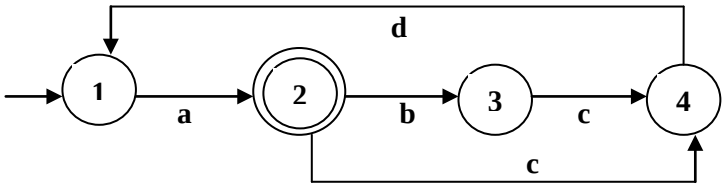
2. Describe the language denoted by the regular expression [b-df-hj-np-tv-z]$^+$. If $\sum$ = set of lower case letters, give a shorter regular expression for the same set of strings.

3. Let $\sum$ = set of lower case letters, write a regular expression for the language of all strings of lower case letters that contain the five vowels in order.

4. Write a regular expression that defines a C-like, fixed-decimal literal with no superfluous leading or trailing zeros. That is, 0.0, 123.01 and 123001.0 are legal, but 00.0, 1.000, 0123.6 and 5.30 are illegal.

5. Let *Seq*(*x, y*) be the set of all strings (of length 1 or more) composed of alternating *x*'s and *y*'s. For example, *Seq*(*a, b*) contains *a, b, ab, ba, aba, bab, abab, baba*, and so on. Write a regular expression that defines *Seq*(*x, y*).

6. Write a regular expression that defines the strings recognized by the DFA below.



7. Write a regular expression that defines strings of 'a's of length 5k+1 (k $\geq$0) and translate it to an NFA. Then transform the NFA to a DFA by subset construction (no need to optimize the DFA).

8. Translate the regular expressions $a|(bc)^*d$ to an NFA. Then transform the NFA to a DFA by subset construction (no need to optimize the DFA).

Questions not covered in tutorial class

A. Write a regular expression that defines a comment delimited by ## markers, which allows single #'s within the comment body.

B. Write a regular expression for an octal integer literal. An octal integer literal is an octal digit, or a non-zero octal digit followed by one or more octal digits.

C. Write a regular expression for an unsigned number (integer or floating point) such as 580, 12.01230, 67.506E4, 09.0E-10 or 59E2. Note that superfluous leading or trailing zeros are allowed.

D. Transform the following NFA to DFA with subset construction.