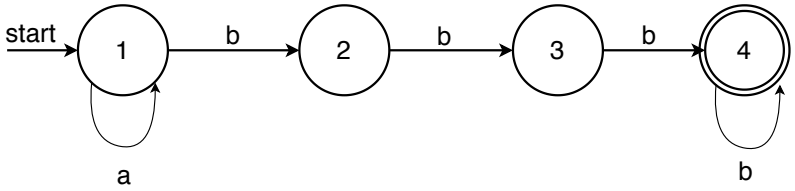
Tutorial 2 : Lexical Analysis

Đỗ Đăng Khôi - 1711807
khoi.do.1711807@hcmut.edu.vn

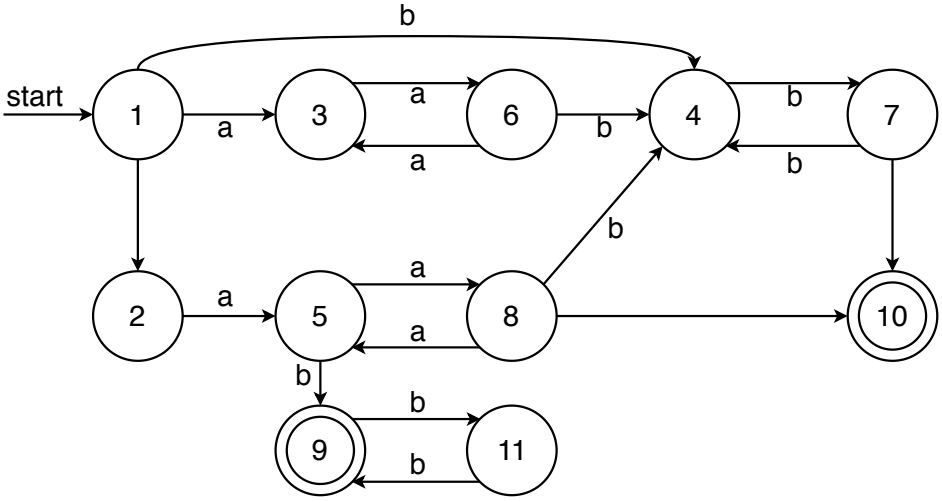
September 2019

4. Find regular expressions and state diagrams of the equivalent NFA for each of the following descriptions.

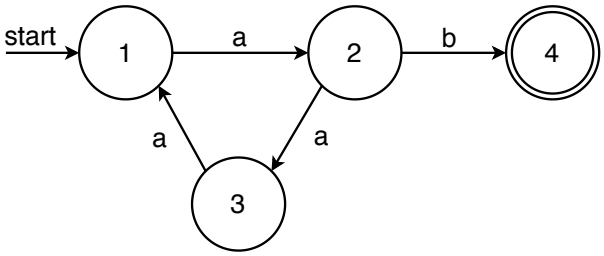
a. $a^n b^m \mid n \geq 0, m > 2$, RegExp : **a*bbb+**



b. $a^n b^m \mid n + m > 0, n + m$ is even, RegExp : **(aa)*(bb)+ | (aa)+(bb)* | a(aa)*b(bb)***



c. $a^n b \mid n \mod 3 = 1$, RegExp : **a(aaa)*b**



1. Use ANTLR to write regular expressions describing a Pascal identifier that must begin with a lowercase letter ('a' to 'z'), but may continue with many characters which are lowercase letter or digit ('0' to '9').

```
ID: [a-z]([a-z]|[0-9])*;
```

2. Use fragment in ANTLR to rewrite the regular expression for the above token Identifier.

```
fragment Digit   : [0-9];
fragment Lowcase: [a-z];

ID: Lowcase(Lowcase|Digit)*;
```

3. Use ANTLR to write regular expressions describing the following Pascal tokens:

a. For a number to be taken as "real" (or "floating point") format, it must either have a decimal point, or use scientific notation. For example, 1.0, 1e-12, 1.0e-12, 0.000000001 are all valid reals. At least one digit must exist on either side of a decimal point.

```
fragment Digit: [0-9];
fragment Negative: '-';
fragment Dot: '.';
fragment Exponent: [eE] Negative? Digit+;
FLOATLIT
    : Digit+ Dot (Digit)* Exponent?
    | Digit* Dot (Digit)+ Exponent?
    | Digit+ Exponent
    ;
```

b. Strings are made up of a sequence of characters between single quotes: 'string'. The single quote itself can appear as two single quotes back to back in a string: 'isn''t'.

```
fragment SingleQuote: '\'';
fragment DoubleQuote: '\'\''
STRLIT : SingleQuote (DoubleQuote | ~('\''))* SingleQuote;
```