CS3012 Formal Languages

Exercises 3: Regular Expressions and Finite State Automata

1. Give regular expressions which exactly define the following languages. The alphabet in each case is given at the end.
    - (i).   contains exactly one a, besides bs. $\{a, b\}$
    - (ii).   no a appears without another a beside it. $\{a, b\}$
    - (iii).   has 00 or 11 as a substring. $\{0, 1\}$
    - (iv).   has no b occurring anywhere after any a. $\{a, b, c\}$
    - (v).   contains exactly two as or exactly two bs. $\{a, b\}$
    - (vi).   has a substring abc or bc. $\{a, b, c\}$
    - (vii). $\{a, b\}^* - \{a\}$ $\{a,b\}$ [i.e. any string of as and bs, but not the string "a"]
    - (viii). contains an even number of as and an even number of bs. $\{a, b\}$
    - (ix).   represents a number divisible by 3. $\{0, 1\}$

Answer:
(i) $b^* a b^*$
- (ii) $(b + aaa^*)^*$
- (iii) $(0 + 1)^*(00 + 11)(0 + 1)^*$
- (iv) $(b + c)^*(a + c)^*$
- (v) $b^* a b^* a b^* + a^* b a^* b a^*$
- (vi) $(a + b + c)^* bc(a + b + c)^*$
- (vii) $\lambda + a(a + b)^+ + b(a + b)^*$
- (viii) $(aa + bb + (ab + ba)(aa + bb)^*(ba + ab))^*$
- (ix) $(0 + 1(01^*0)^*1)^*$

also $(0+11+10(1+00)\text{*}01)\text{*}$

---

2. Describe in plain English the languages defined by the following regular expressions
    - (i) $(a + b)^*$
    - (ii) $(a + b)^* ab(a + b)^*$
    - (iii) $a^2 + b^2$
    - (iv) $((a + b)^2)^*$
    - (v) $a^* b$
    - (vi) $1(0 + 1)^*$
    - (vii) $(0 + 1)^* 011$
    - (viii) $011^* + 100^*$
    - (ix) $(00 + 1)^*$
    - (x) $0 (1+0)^* 0$

Answer:

- (i)   strings of a's and b's
- (ii)   strings of a's and b's, with a substring ab.
- (iii) aa or bb.
- (iv)   strings of even length.
- (v)   strings of a's followed by a b.
- (vi)   non-zero binary numbers
- (vii)   strings of 0's and 1's ending in 011.
- (viii) 0 followed by a non-zero number of 1's, or 1 followed by a non-zero number of 0's
- (ix)   strings of 0's and 1's in which 0's only occur in even length substrings.
- (x)   strings of 0's and 1's which cannot begin or end with a 1.

3. A certain programming language allows real constants to be written in exponent form - e.g. 3.25E6 (for 3.25x106) or 2.16E-5 (for 2.16x10-5). There must be exactly one digit before the decimal point (preceded by an optional sign), at least one digit after the decimal point, and the exponent can be a positive or negative integer. Write a regular expression specifying the language of such numbers. Define any abbreviations you use.

Answer:
$S = \lambda + $ "+" + "-" $\quad\quad D = $ "." $\quad\quad N = 0 + 1 + 2 + ... + 9$
$\quad\quad SNDN^+ESN^+$

4. Give a regular expression for all character strings which contain the word "cat". Assume only letters and spaces occur in the strings, and that words are separated by single spaces.
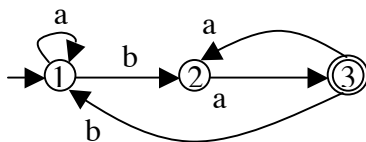
Answer:
Let $X = a + b + ... + z + A + ... + Z$ Let # be a space Regular expression $= (X^+\#)^* cat(\#X^+)^*$

5. Convert the FSAs below to regular expressions, using the FSA -> RegExp algorithm (or by inspection).

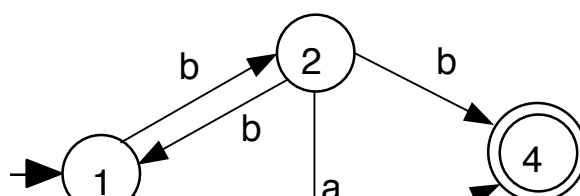| Answer: |
| :--- |
| Add states i and f
Remove 2:
Add (1,ba,3)(3,aa,3)
Remove 3:
Add (1,ba(aa)*,f)(1,ba(aa)*b,1)
Remove 1:
$(i,(a+ba(aa)*b)*ba(aa)*,f)$
you can get different correct answers depending on the order you remove states in |

Answer:

```
        bb
         ⟲
              bb
   →  (1) ─────────────────→ ((4))
         \
          \ b+ba          a
           ↘              ↗
            (3) ─────────
              ⟲
               b

        bb
         ⟲
              bb+((b+ba)b*a)
   →  (1) ─────────────────→ ((4))
```

```
                    (bb)*(bb+((b+ba)b*a))
   →  (i) ───────────────────────────────────→ ((f))
```