Regular Expression:

"Regular expression (R.E.) is an algebraic way (expression language) to denote regular language"

(1) $\forall a \in \Sigma$, a is a R.E. $\Rightarrow L(a) = \{a\}$

ϵ is a R.E $\Rightarrow L(\epsilon) = \{\epsilon\}$

ϕ is a R.E $\Rightarrow L(\phi) = \phi$

$\longrightarrow$ Primitive regular expressions.

(2) If r_1 and r_2 are R.E, then so are $\underbrace{r_1 + r_2,\ r_1 \cdot r_2,\ r_1^*}_{\text{operators that combine R.E's}}$ and $\underbrace{(r_1)}_{\text{Disambiguate notation}}$

$\Rightarrow L(r_1 + r_2) = L(r_1) \cup L(r_2)$,

$L(r_1 \cdot r_2) = L(r_1) \cdot L(r_2)$

$L(r_1^*) = (L(r_1))^*$.

(3) A string s is a regular expression if and only if it can be obtained by applying operation in (2) to the primitive regular expression in (1)

Ex: $\Sigma = \{a, b\}$

$(a+\epsilon)\cdot(b\cdot a+\phi^*)$ is a valid. regular expression,

but $(a \cap \epsilon) - (b^R \cdot \phi)$ is NOT a valid regular expression.

Order of Precedence (excluding ( ))

Kleene's closure
Positive closure
Concatenation
Union.

Examples: let $\Sigma = \{a, b\}$.

1. starts with ab $\quad ab(a+b)^*$
2. Ending with ab. $\quad (a+b)^*ab$
3. Contains a substring aab $\quad (a+b)^*aab(a+b)^*$
4. start and end with a $\quad a + a(a+b)^*a$
5. start and end with the same symbol.

$$a(a+b)^*a + b(a+b)^*b + a + b + \epsilon$$

⑥ start and end with different symbol

$$a(a+b)^*b + b(a+b)^*a$$

⑦ $|w| = 3$ $\quad (a+b)(a+b)(a+b)$

⑧ $|w| \geq 3$ $\quad (a+b)(a+b)(a+b)(a+b)^*$

⑨ $|w| \leq 3$ $\quad (a+b+\epsilon)(a+b+\epsilon)(a+b+\epsilon)$

⑩ $|w|_a = 2$ $\quad b^*ab^*ab^*$.

⑪ $|w| = 0 \pmod 3$ $\quad ((a+b)(a+b)(a+b))^*$

Important identities for regular expressions:

I_1: $\phi + R = R$

I_2: $\phi R = R\phi = \phi$

I_3: $\Lambda R = R\Lambda = R$

I_4: $\Lambda^* = \Lambda$, $\phi^* = \Lambda$

I_5: $R + R = R$

I_6: $R^*R^* = R^*$.

I_7: $RR^* = R^*R$

I_8: $(R^*)^* = R^*$.

I_9: $\Lambda + RR^* = R^*$

I_{10}: $(PQ)^*P = P(QP)^*$.

I_{11}: $(P+Q)^* = (P^*Q^*)^* = (P^*+Q^*)^*$

let P and Q be two regular expressions over Σ, if P does not contain ϵ, then the following equation in R

$$R = Q + RP$$

has a unique solution $R = QP^*$

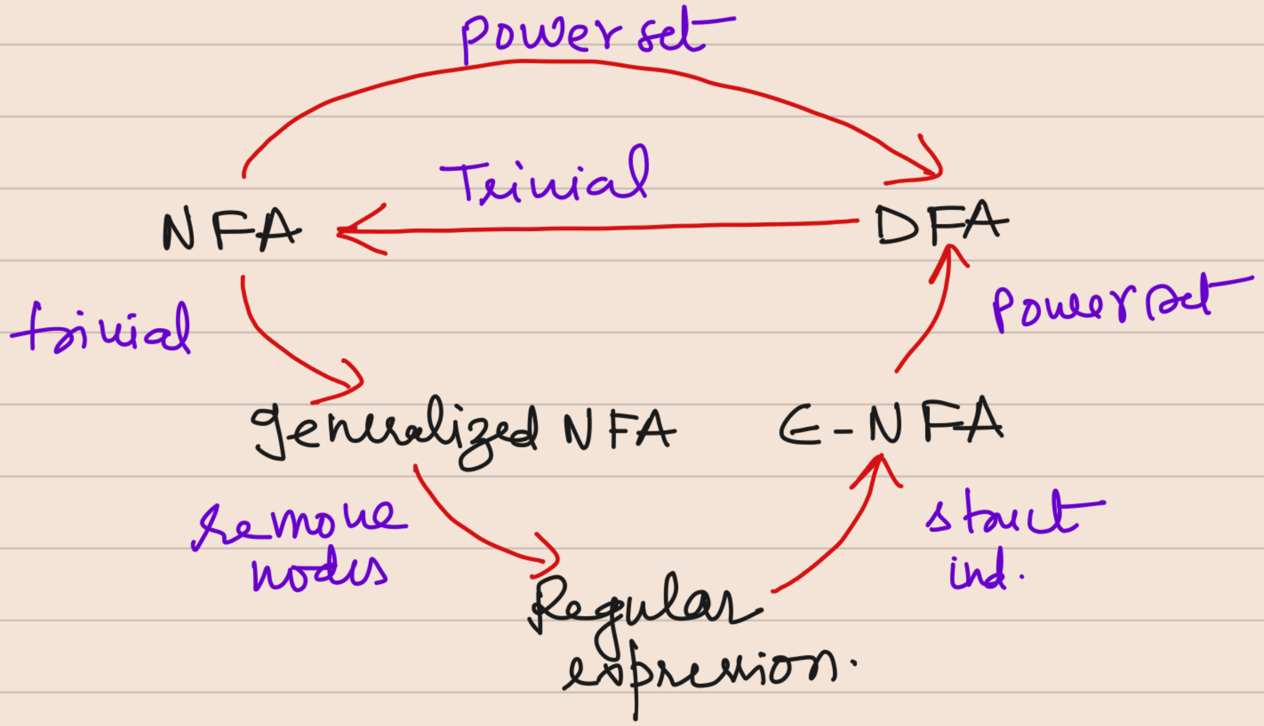
Note:

$$R = PR + Q \qquad R = P^*Q \;\} \text{ Right}$$

$$R = RP + Q \qquad R = QP^* \;\} \text{ left}$$

Kleene's theorem:

"the set of regular languages, the set of NFA-recognizable languages and the set of DFA-recognizable languages are all same"

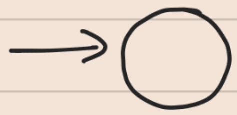


Converting regular expressions to DFA :

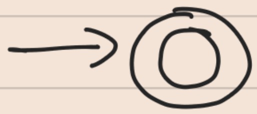
Note: To convert a R.E to an NFA, first convert it to an ϵ-NFA, then convert that to a DFA.

R.E

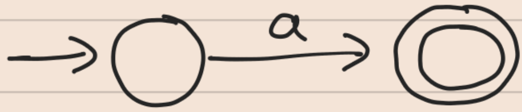
$r = \phi$



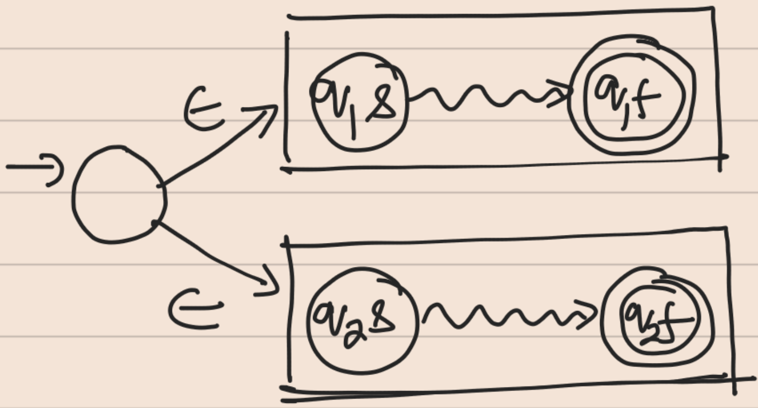
$r = \epsilon$



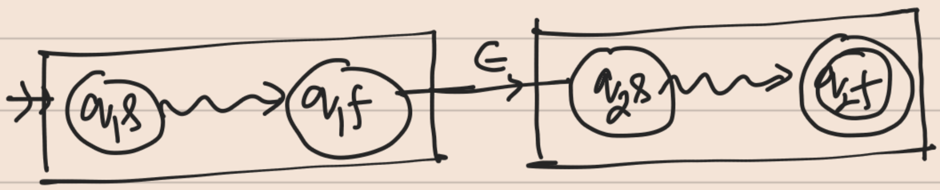
$r = a$



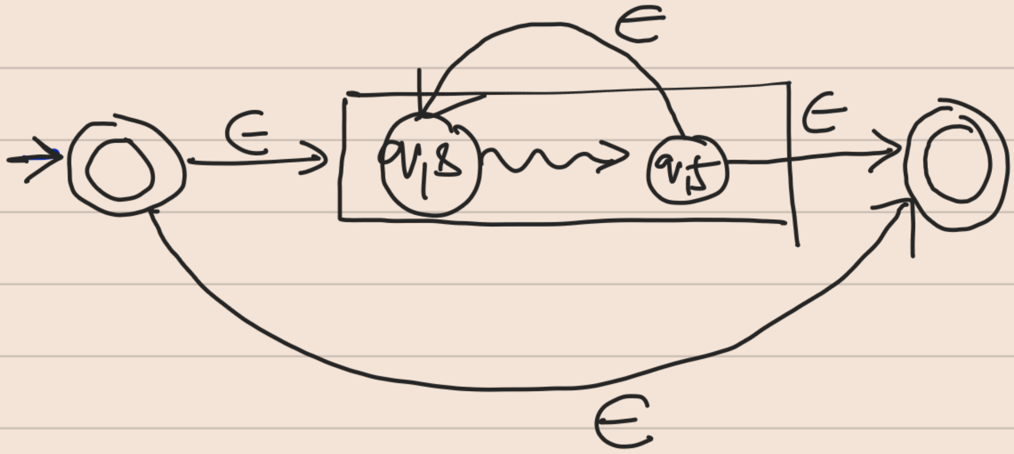
$r = r_1 | r_2$



$r = r_1 \cdot r_2$



$r = r_1^*$

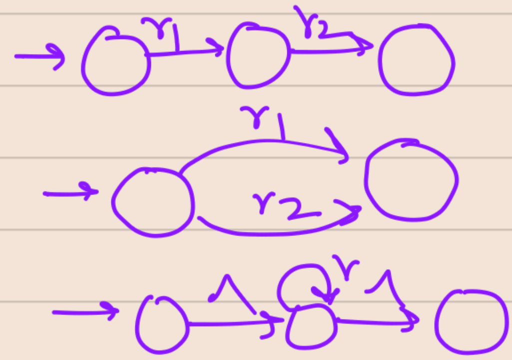


A simplified view: (useful when converting R.E. to DFA)

$r = r_1 \cdot r_2$

$r = r_1 + r_2$

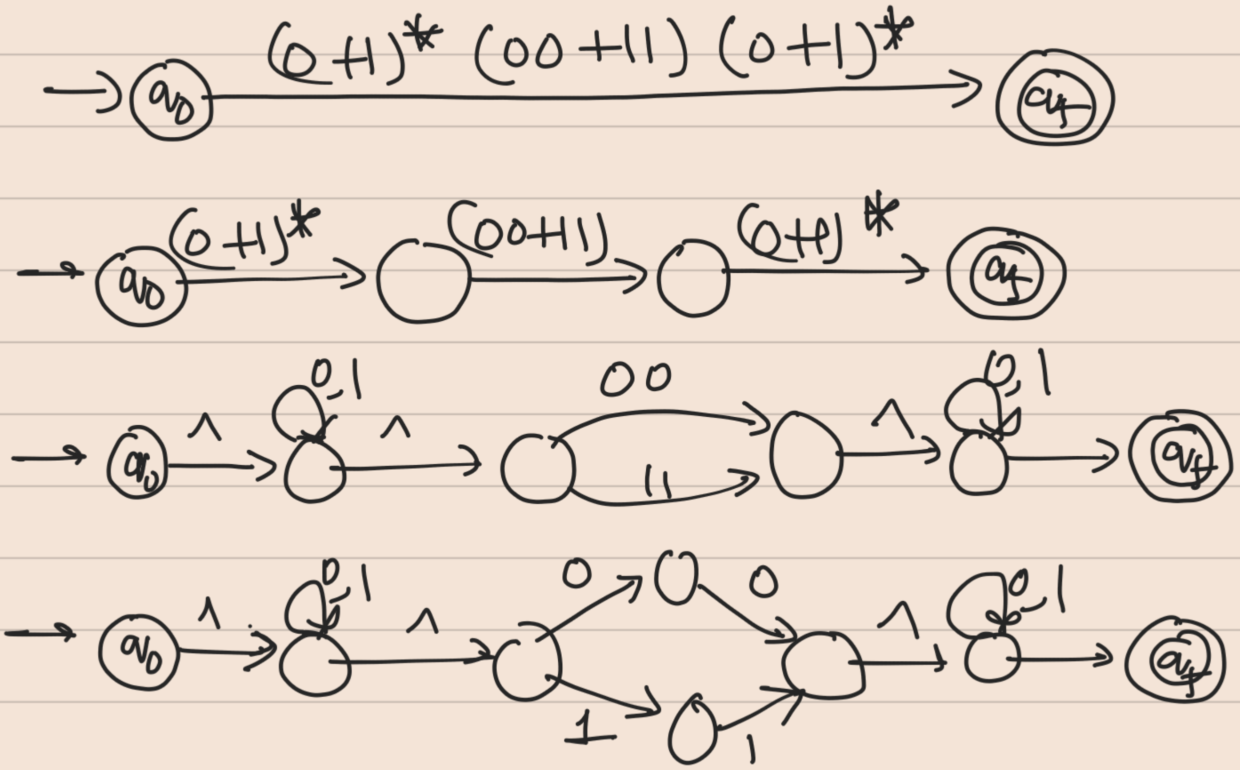
$r = r_1^*$

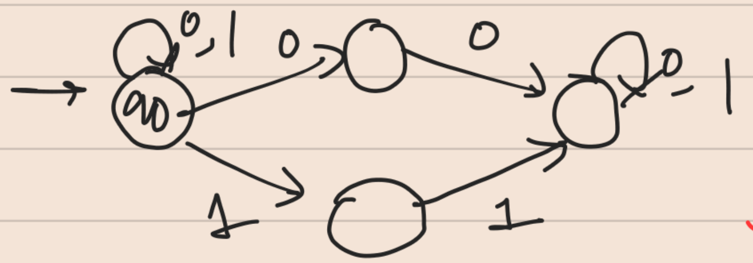


→ Not precise, but useful for solving problems

Ex: Construct a Finite automata equivalent to the regular expression

$$(0+1)^*(00+11)(0+1)^*$$





this is NFA, convert this into a DFA