WEEK 4: Lecture Notes

Regular Expressions:

- $\Sigma \rightarrow$ a finite set of symbols

  $L_1, L_2, L \rightarrow$ language over Σ
  $\subseteq \Sigma^*$

- Concatenation of L_1 and L_2

$$L_1 L_2 = \{ xy \mid x \in L_1, y \in L_2 \}$$

$$L^0 = \{ \varepsilon \}$$

$$L^i = L L^{i-1} \quad \text{for } i \geq 1$$

- Kleene Closure of L

$$L^* = \bigcup_{i=0}^{\infty} L^i$$

- Positive Closure

$$L^+ = \bigcup_{i=1}^{\infty} L^i \rightarrow \text{contains } \varepsilon \text{ iff } L \text{ does}$$

Example:

$$L_1 = \{10, 1\}, \quad L_2 = \{011, 11\}$$

$$L_1 L_2 = \{10011, 1011, 1011, 111\}$$

Example:

$$L = \{10, 11\}$$

$$L^* = \{ \varepsilon, 10, 11, 1010, 1011, 1110, 1111, \ldots \}$$

- $\Sigma \rightarrow$ an alphabet

The regular expressions over Σ and the sets (languages) that they denote are defined recursively as:

i) ϕ is a regular expression and denotes the empty set

ii) ε is a regular expression and denotes the set $\{\varepsilon\}$

iii) for each $a \in \Sigma$, a is a regular expression and denotes the set $\{a\}$

iv) if r and s are regular expressions denoting the language R and S respectively, then $(r+s)$, (rs) and (r^*) are regular expressions that denote the sets $R \cup S$, RS and R^* respectively.

Precedence of Regular Expression operations

- $*$ has higher precedence than concatenation or $+$
- concatenation has higher precedence than $+$

Example: $((0(1^*))+0) \rightarrow 01^*+0$

rr^* is same as r^+

Notation:

- $r \rightarrow$ a regular expression
- $L(r) \rightarrow$ the set / language denoted by r

- $L(\phi) = \phi$, $L(\varepsilon) = \varepsilon$, $L(a) = \{a\}$
- $L(r+s) = L(r) \cup L(s)$, $L(rs) = L(r)\, L(s)$.

Example:

Consider the language consisting of strings of a's and b's containing aab

$\rightarrow \quad (a+b)^* \, aab \, (a+b)^*$

Example:

Set of all strings of 0's and 1's with atleast two consecutive 0's $\quad (0+1)^* \, 00 \, (0+1)^*$

Example:

$$L(00) = \{00\}, \; L\left((0+1)^*\right) = \{0,1\}^*$$

Example:

$$(1+10)^* = \{\varepsilon, 1, 10, 11, 110, 101, 1010, \ldots\}$$

all binary strings begining with 1's and not having two consecutive 0's

$(1+10)^i$: binary strings begining with 1, not having two consecutive 0's, having i number of 1's.

1101011 : $1 - 10 - 10 - 1 - 1 \in (1+10)^5$

$(0+\varepsilon)(1+10)^*$: all binary strings that do not have two consecutive 0's.

$(0+1)^* \, 011$: all binary strings ending in 011

$0^* 1^* 2^*$: any number of 0's, followed by any number of 1's, followed by any number of 2's

$00^* \, 11^* \, 22^*$: string in $0^*1^*2^*$ with atleast one of each symbol $\rightarrow \; 0^+ 1^+ 2^+$

Algebraic Laws for regular expressions

- commutativity for union $r+s = s+r$
- associativity for union $(r_1+r_2)+r_3 = r_1+(r_2+r_3)$
- associativity for concatenation

  $(r_1 r_2) r_3 = r_1 (r_2 r_3)$

- NOT commutative for concatenation

  $r_1 r_2 \neq r_2 r_1$

  as $01 \neq 10$, where $r_1 = 0$,
  $r_2 = 1$.

- $\phi + r = r + \phi = r$ (ϕ is the identity for union)
- $\varepsilon r = r \varepsilon = r$ (ε is the identity for concatenation)
- $\phi r = r \phi = \phi$ (ϕ is the annihilator for concatenation)

- $\varepsilon + r \neq r$ unless r contains ε

- Distributive laws of concatenation over union

  $r_1(r_2+r_3) = r_1 r_2 + r_1 r_3$

  $(r_1+r_2) r_3 = r_1 r_3 + r_2 r_3$

- Idempotent Law

  $r + r = r$

Laws involving closures

- $(r^*)^* = r$
- $\phi^* = \varepsilon$
- $\varepsilon^* = \varepsilon$
- $r^+ = rr^* = r^*r \rightarrow r^+ = r + rr + rrr + \cdots$

  $r^* = \varepsilon + r + rr + \cdots$

  $\therefore rr^* = r^+$

- $r^* = r^* + \varepsilon$

Other Laws:

- $(r+s)^* = (r^* s^*)^*$ as $(a+b)^*$: set of all strings of a's and b's

  $(a^* b^*)^*$: set of all strings of a's and b's

- $r^* = r^* r^*$ as a^* : set of all strings of a's

  $a^* a^*$: set of all strings of a's

- $r_1 + r_2 r_1 \neq (r_1 + r_2) r_1$

  as $a + ba$ and $(a+b)a$ are different regular expressions

  e.g. string aa is not in $L(a+ba)$ but it is in $L((a+b)a)$

Equivalence of ε-NFA and regular expressions

Theorem: Let r be a regular expression. Then $\exists$ an ε-NFA that accepts $L(r)$.

Proof:

By induction on the number of operators in r, we will show that there is an ε-NFA M, having one final state and no transition out of this final state, s.t. $L(r) = L(M)$

Base: (zero operators)

$r = \varepsilon$ | $r = \phi$ | $r = \{a\}$, $a \in \Sigma$

Induction: (one or more operators)

Assume that the theorem is true for regular expressions with fewer than i operators, $i \geq 1$.

Let r have i operations.

<u>Case I:</u>

$r = r_1 + r_2$, r_1, r_2 must have fewer than i operations

$\therefore \exists$ ε-NFA's M_1, M_2 s.t. $L(r_1) = L(M_1)$
$L(r_2) = L(M_2)$

Let $M_1 = \{ Q_1, \Sigma_1 \cup \{\varepsilon\}, \delta_1, q_1, \{f_1\} \}$

$M_2 = \{ Q_2, \Sigma_2 \cup \{\varepsilon\}, \delta_2, q_2, \{f_2\} \}$

where $Q_1 \cap Q_2 = \phi$

Construct an ε-NFA

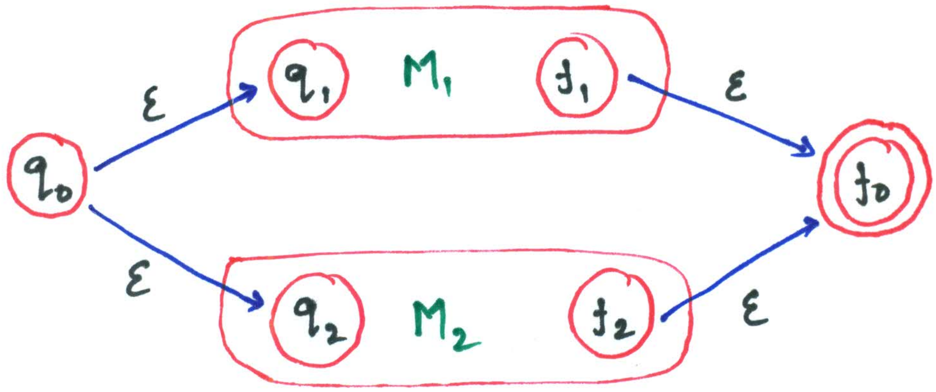
$M = (Q_1 \cup Q_2 \cup \{q_0, f_0\}, \Sigma_1 \cup \Sigma_2 \cup \{\varepsilon\}, \delta, q_0, \{f_0\})$

where δ is defined by

q_0: new initial state

f_0: new final state

- $\delta(q_0, \varepsilon) = \{q_1, q_2\}$
- $\delta(q, a) = \delta_1(q, a)$ for $q \in Q_1 - \{f_1\}$, $a \in \Sigma_1 \cup \{\varepsilon\}$
- $\delta(q, a) = \delta_2(q, a)$ for $q \in Q_2 - \{f_2\}$, $a \in \Sigma_2 \cup \{\varepsilon\}$
- $\delta(f_1, \varepsilon) = \delta(f_2, \varepsilon) = f_0$

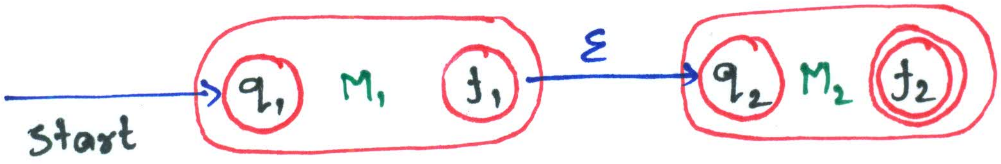


$q_0 \rightsquigarrow f_0$: path labeled x in M iff either $q_1 \rightsquigarrow f_1$: path labeled x in M_1 or $q_2 \rightsquigarrow f_2$ path labeled x in M_2.

Case II

$r = r_1 r_2$

$M = (Q_1 \cup Q_2,\ \Sigma_1 \cup \Sigma_2 \cup \{\varepsilon\},\ \delta,\ \{q_1\}, \{f_2\})$



where δ is defined by

- $\delta(q, a) = \delta_1(q, a)$ for $q \in Q_1 - \{f_1\}$ and $a \in \Sigma_1 \cup \{\varepsilon\}$
- $\delta(f_1, \varepsilon) = \{q_2\}$
- $\delta(q, a) = \delta_2(q, a)$ for $q \in Q_2$ and $a \in \Sigma_2 \cup \{\varepsilon\}$

$q_1 \rightsquigarrow f_2$ in M

iff

$q_1 \overset{x}{\sim} f_1 \xrightarrow{\varepsilon} q_2 \overset{y}{\sim} f_2$

$L(M) = \{xy \mid x \in L(M_1),\ y \in L(M_2)\} = L(M_1)L(M_2)$

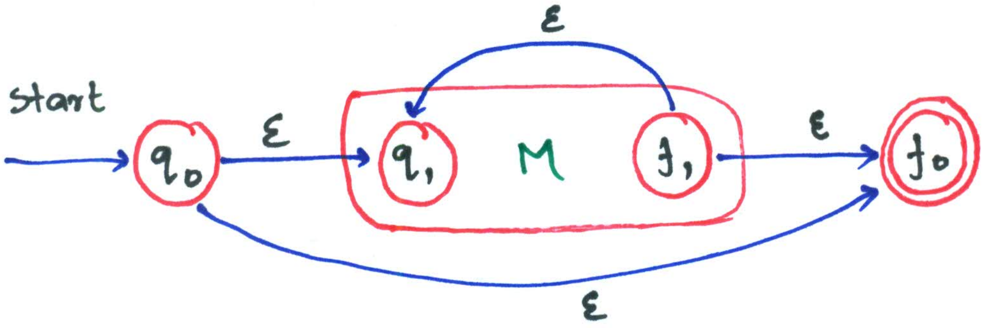
Case III

$r = r_1^*$

$L(M_1) = L(r_1)$ as r has fewer than i operators.

Construct

$$M = \{Q \cup \{q_0, f_0\}, \Sigma \cup \{\varepsilon\}, \delta, q_0, f_0\}$$



δ is defined by

- $\delta(q_0, \varepsilon) = \{q_1, f_0\} = \delta(f_1, \varepsilon)$
- $\delta(q_0, a) = \delta_1(q, a)$ for $q \in Q - \{f_1\}$ and $a \in \Sigma_1 \cup \{\varepsilon\}$

$q_0 \overset{x}{\rightsquigarrow} f_0$ in M iff

$$q_0 \xrightarrow{\varepsilon} f_0$$

or

$$q_0 \xrightarrow{\varepsilon} q_1 \overset{x_1}{\rightsquigarrow} f_1 \xrightarrow{\varepsilon} q_1 \overset{x_2}{\rightsquigarrow} f_1 \xrightarrow{\varepsilon} q_1 \overset{x_i}{\rightsquigarrow} \dots f_1 \xrightarrow{\varepsilon} f_0$$

where $x_1, x_2, \dots x_i \in L(M_1)$

and

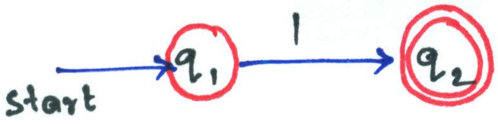
$x = x_1 x_2 \dots x_i$ for some $i \geq 0$

($i = 0$ means $x = \varepsilon$)

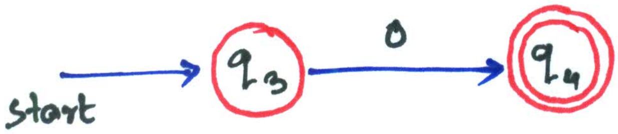
$$\therefore \; L(M) = (L(M_1))^*$$

Example:

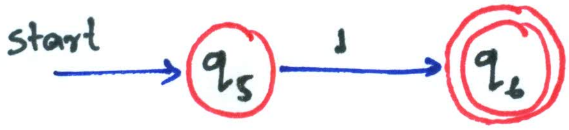
- $r = 01^* + 1 = r_1 + r_2$; $r_1 = 01^*$, $r_2 = 1$



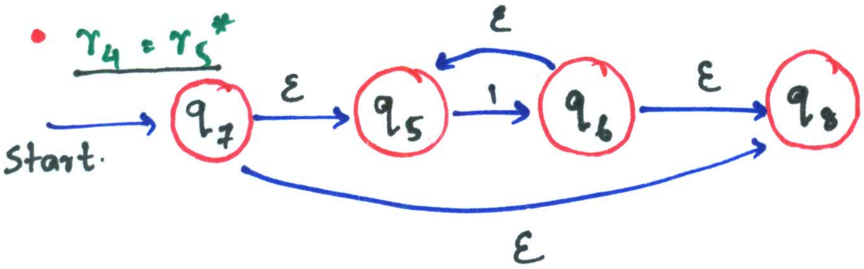
- $r_1 = 01^* = r_3 r_4$, $r_3 = 0$, $r_4 = 1^*$



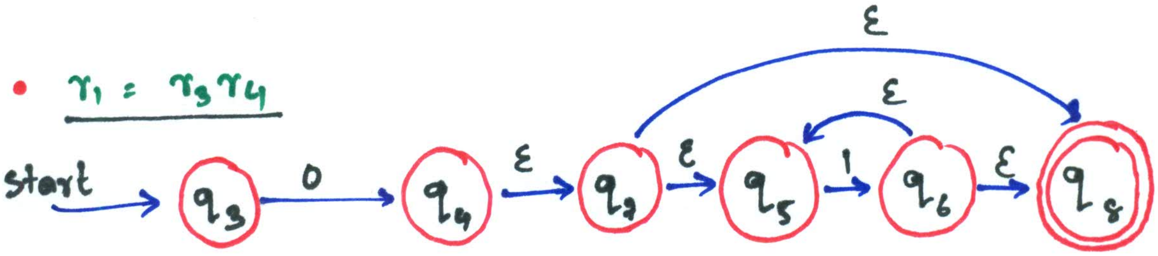
- $r_4 = 1^* = r_5^*$, $r_5 = 1$



- $r_4 = r_5^*$



- $r_1 = r_3 r_4$



- $r = r_1 + r_2$