Course 3

Regular Languages

Why?

1. Search engine – succes of Google
2. Unix commands
3. Programming languages – new feature

Remember

- **Grammar**

$G=(N,\Sigma,P,S)$

$L(G)=\{w\in\Sigma^* \mid S \Rightarrow w\}$

- Finite automaton

$M = (Q,\Sigma,\delta,q_0,F)$

$L(M)=\{ w \in \Sigma^* \mid (q_0,w) \vdash (q_f ,\varepsilon) , q_f \in F \}$

Remember: Regular grammars

- $G = (N, \Sigma, P, S)$ right linear grammar if

  $\forall p \in P$: $A \rightarrow aB$ or $A \rightarrow b$, where $A,B \in N$ and $a,b \in \Sigma$

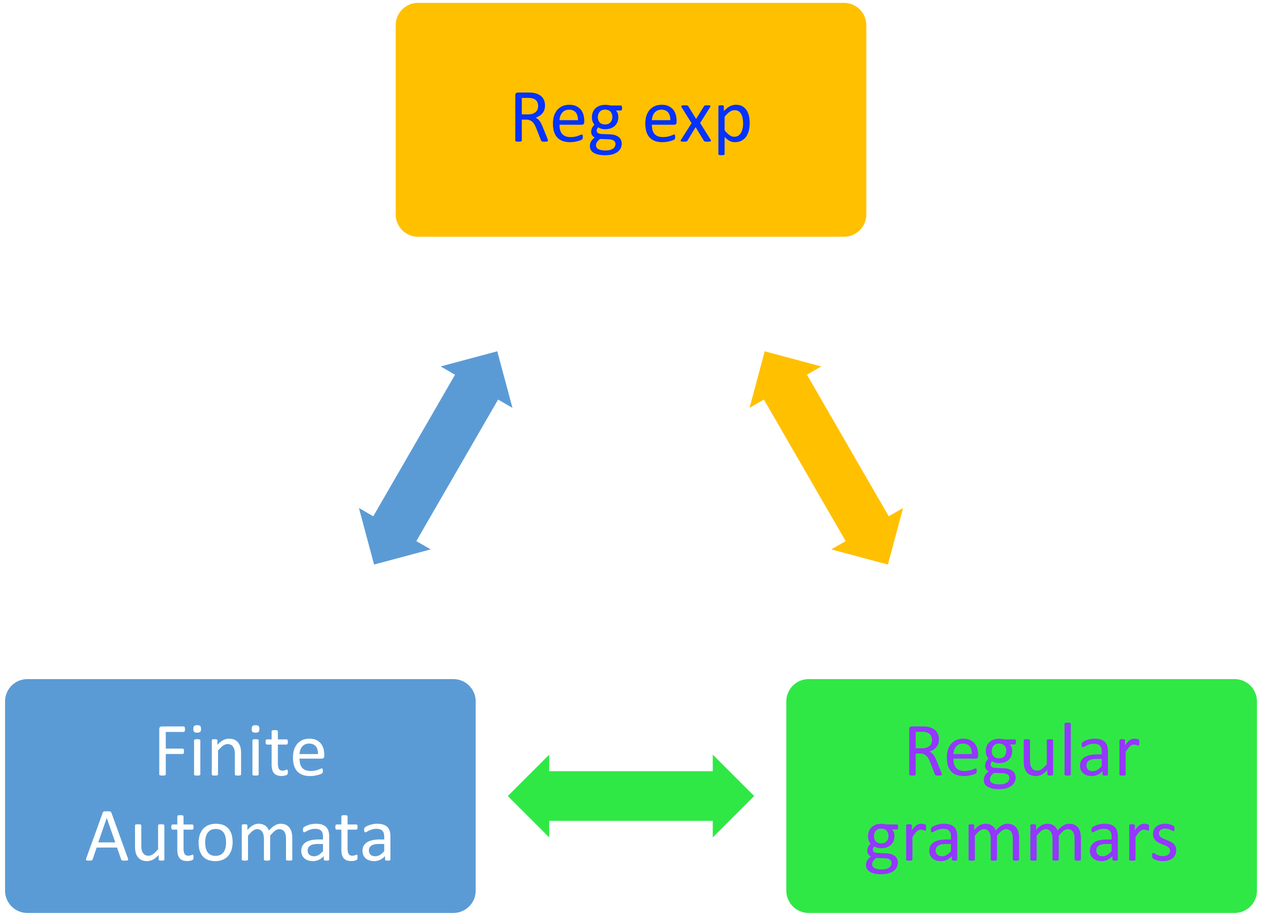
- $G = (N, \Sigma, P, S)$ regular grammar if
  - G is right linear grammar

  and

  - $A \rightarrow \varepsilon \notin P$, with the exception that $S \rightarrow \varepsilon \in P$, in which case S does not appear in the rhs (right hand side) of any other production

- $L(G) = \{w \in \Sigma^* \mid S \overset{*}{\Rightarrow} w\}$ - right linear language

Regular languages / sets



Theorem 1: For any regular grammar $G=(N, \Sigma, P, S)$ there exists a FA $M=(Q, \Sigma, \delta, q_0, F)$ such that $L(G) = L(M)$

Proof: construct M based on G

$Q = N \cup \{K\}, K \notin N$

$q_0 = S$

$F = \{K\} \cup \{S \mid \text{if } S \to \varepsilon \in P\}$

δ: if $A \to aB \in P$ then $\delta(A,a) = B$

if $A \to a \in P$ then $\delta(A,a) = K$

Prove that L(G) = L(M) *($w \in L(G) \Leftrightarrow w \in L(M)$):*

$S \overset{*}{\Rightarrow} w \Leftrightarrow (S,w) \overset{*}{\vdash} (qf, \varepsilon)$

$w = \varepsilon$: $S \overset{*}{\Rightarrow} \varepsilon \Leftrightarrow (S, \varepsilon) \overset{*}{\vdash} (S, \varepsilon)$ – true

$w=a_1a_2...a_n$: $S \overset{*}{\Rightarrow} w \Leftrightarrow (S,w) \overset{*}{\vdash} (K, \varepsilon)$

$S \Rightarrow a_1A_1 \Rightarrow a_1a_2A_2 \Rightarrow ... \Rightarrow a_1a_2...a_{n-1}A_{n-1} \Rightarrow a_1a_2...a_{n-1}a_n$

$S \Rightarrow a_1A_1$ exists if $S \to a_1A_1$ and then $\delta(S,a_1)=A_1$

$A_1 \to a_2A_2$: $\delta(A_1,a_2)=A_2$ . . .

$A_{n-1} \to a_n$: $\delta(A_{n-1},a_n)=K$

$(S,a_1a_2...a_n) \vdash (A_1,a_2...a_n) \vdash (A_2,a_3...a_n) \vdash ... \vdash (A_{n-1},a_n) \vdash (K, \varepsilon)$, $K \in F$

Theorem 2: For any FA $M=(Q, \Sigma, \delta, q_0, F)$ there exists a regular grammar $G=(N, \Sigma, P, S)$ such that $L(G) = L(M)$

Proof: construct G based on M

$N = Q$

$S = q_0$

P: if $\delta(q,a) = p$ then $q \rightarrow ap \in P$

if $p \in F$ then $q \rightarrow a \in P$

if $q_0 \in F$ then $S \rightarrow \varepsilon$

Prove that L(M) = L(G) *($w \in L(M) \Leftrightarrow w \in L(G)$):*

$P(i)$: $q \overset{i+1}{\Rightarrow} x \Leftrightarrow (q,x) \overset{i}{\vdash} (q_f, \varepsilon)$, $q_f \in F$ -prove by induction

Apply P : $q_0 \overset{i+1}{\Rightarrow} w \Leftrightarrow (q_0,w) \overset{i}{\vdash} (q_f, \varepsilon)$, $q_f \in F$

If i=0: $q \Rightarrow x \Leftrightarrow (q,x) \overset{0}{\vdash} (q_f, \varepsilon)$ $(x = \varepsilon, q = q_f)$ $q \Rightarrow \varepsilon \Leftrightarrow q_0 \rightarrow \varepsilon$, $q_0 \in F$

Assume $\forall$ $k \le i$ P is true

$q \overset{i+1}{\Rightarrow} x \Leftrightarrow (q,x) \overset{i}{\vdash} (q_f, \varepsilon)$

For $q \in N$ apply "$\Rightarrow$" : $q \Rightarrow ap \overset{i}{\Rightarrow} ax$

If $q \Rightarrow ap$ then $\delta(q,a) = p$; if $p \overset{i}{\Rightarrow} ax$ then $(p,x) \overset{i-1}{\vdash} (q_f, \varepsilon)$, $qf \in F$

THEN $(q,ax) \overset{i}{\vdash} (q_f, \varepsilon)$, $qf \in F$

Regular sets

Definition: Let $\mathbf{\Sigma}$ be a finite alphabet. We define <u>regular sets</u> over $\mathbf{\Sigma}$ recursively in the following way:

1. $\boldsymbol{\Phi}$ is a regular set over $\mathbf{\Sigma}$ (empty set)
2. $\{\boldsymbol{\varepsilon}\}$ is a regular set over $\mathbf{\Sigma}$
3. $\{a\}$ is a regular set over $\mathbf{\Sigma}$, $\forall\ a \in \mathbf{\Sigma}$
4. If P, Q are regular sets over $\mathbf{\Sigma}$, then $P \cup Q$, PQ, P^* are regular sets over $\mathbf{\Sigma}$
5. Nothing else is a regular set over $\mathbf{\Sigma}$

Regular expressions

Definition: Let $\boldsymbol{\Sigma}$ be a finite alphabet. We define <u>regular expressions</u> over $\boldsymbol{\Sigma}$ recursively in the following way:

1. $\boldsymbol{\Phi}$ is a regular expression denoting the regular set $\boldsymbol{\Phi}$ (empty set)
2. $\boldsymbol{\varepsilon}$ is a regular expression denoting the regular set $\{\boldsymbol{\varepsilon}\}$
3. **a** is a regular expression denoting the regular set {a}, $\forall\ a \in \boldsymbol{\Sigma}$
4. If **p,q** are regular expression denoting the regular sets P, Q then:
   - **p+q** is a regular expression denoting the regular set $P \cup Q$,
   - **pq** is a regular expression denoting the regular set PQ,
   - **p*** is a regular expression denoting the regular set P*
5. Nothing else is a regular expression

Remarks:

Examples

1. $p^+ = pp^*$
2. Use paranthesis to avoid ambiguity
3. Priority of operations: *, concat, + (from high to low)
4. For each regular set we can find at least one regular exp to denote it (there is an infinity of reg exp denoting them)
5. For each regular exp, we can construct the corresponding regular set
6. 2 regular expressions are **equivalent** **iff** they denote the same regular set

Algebraic properties of regular exp

Let $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ be regular expressions.

1. $\boldsymbol{\alpha + \beta = \beta + \alpha}$
2. $\boldsymbol{\Phi^* = \varepsilon}$
3. $\boldsymbol{\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma}$
4. $\boldsymbol{\alpha(\beta\gamma) = (\alpha\beta)\gamma}$
5. $\boldsymbol{\alpha (\beta + \gamma) = \alpha\beta + \alpha\gamma}$
6. $\boldsymbol{(\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma}$
7. $\boldsymbol{\alpha \varepsilon = \varepsilon \alpha = \alpha}$
8. $\boldsymbol{\Phi\alpha = \alpha\Phi = \Phi}$
9. $\boldsymbol{\alpha^* = \alpha + \alpha^*}$
10. $\boldsymbol{(\alpha^*)^* = \alpha^*}$
11. $\boldsymbol{\alpha + \alpha = \alpha}$
12. $\boldsymbol{\alpha + \Phi = \alpha}$

Reg exp equations

- Normal form: **X = aX + b**

  where a,b – reg exp

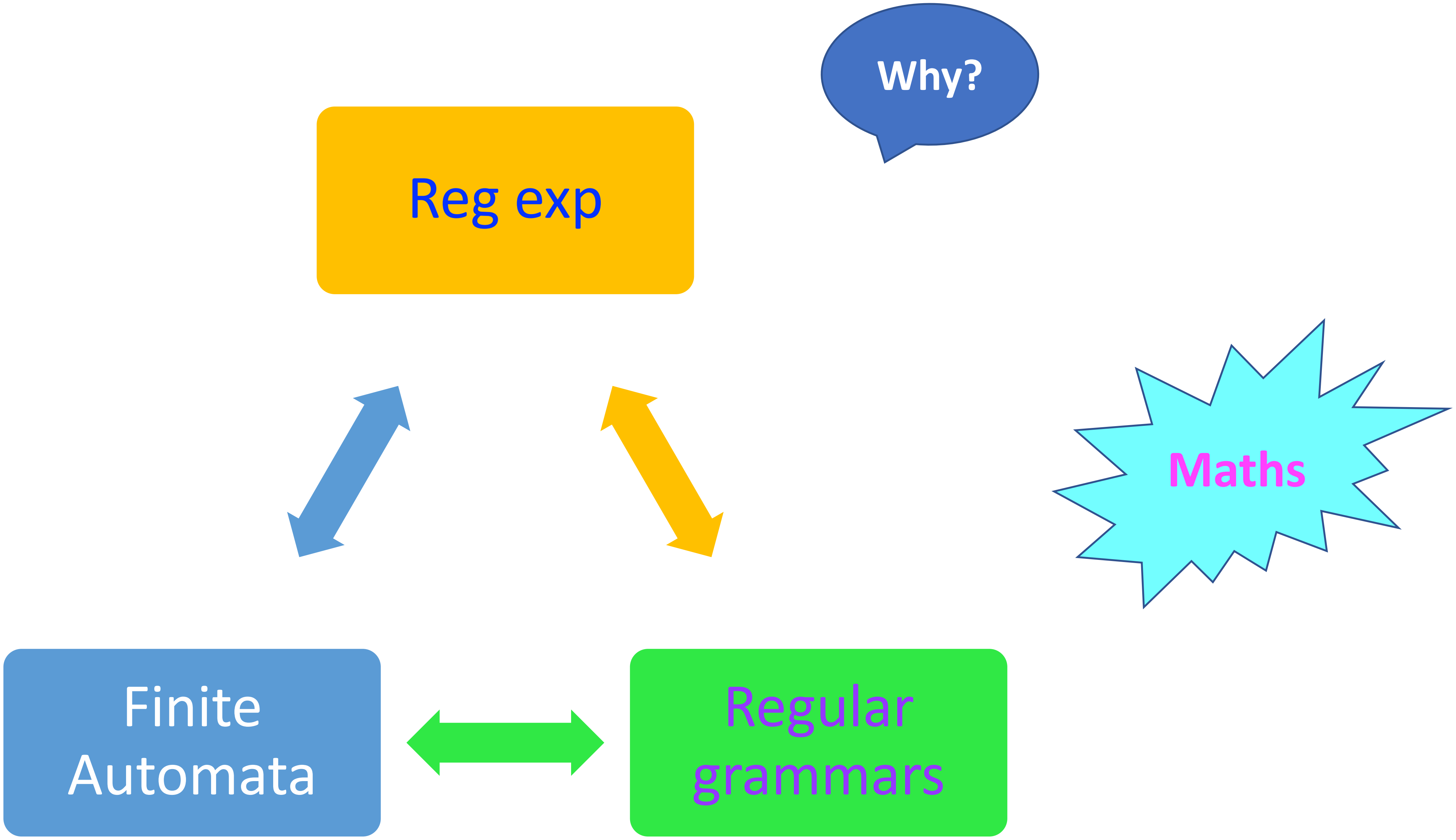
- Solution: **X = a*b**

$\boxed{a\,a^*b + b = (aa^* + \boldsymbol{\varepsilon})b = a^*b}$

- System of reg exp equations:

$$\begin{cases} X = a_1X + a_2Y + a_3 \\ Y = b_1X + b_2Y + b_3 \end{cases}$$

- Solution: Gauss method (substitution)


Why?
Reg exp
Maths
Finite Automata
Regular grammars

Prop: *Regular sets are right linear languages*

Lemma 1: $\boldsymbol{\Phi}$,{$\boldsymbol{\varepsilon}$}, {a},∀a∈$\boldsymbol{\Sigma}$ are right linear languages

Proof: constructive

i. G = ({S}, $\boldsymbol{\Sigma}$, $\boldsymbol{\Phi}$, S) – regular grammar such that L(G) = $\boldsymbol{\Phi}$

ii. G = ({S}, $\boldsymbol{\Sigma}$,{S→$\boldsymbol{\varepsilon}$}, S) – regular grammar such that L(G) ={$\boldsymbol{\varepsilon}$}

iii. G = ({S}, $\boldsymbol{\Sigma}$,{S→a}, S) – regular grammar such that L(G) ={a}

Lemma 2: If L_1 and L_2 are right linear languages then:

$L_1 \cup L_2$, L_1L_2 and $L_1{}^*$ are right linear languages.

Proof: constructive

L_1, L_2 right linear languages => $\exists G_1, G_2$ such that

$G_1 = (N_1, \boldsymbol{\Sigma}, P_1, S_1)$ and $L_1 = L(G_1)$

$G_2 = (N_2, \boldsymbol{\Sigma}, P_2, S_2)$ and $L_2 = L(G_2)$ assume $N_1 \cap N_2 = \emptyset$

i. $G_3 = (N_3, \Sigma, P_3, S_3)$

$N_3 = N_1 \cup N_2 \cup \{S_3\}$

$P_3 = P_1 \cup P_2 \cup \cancel{\{S_3 \rightarrow S_1 \mid S_2\}}$

$\{S_3 \rightarrow \alpha_1 \mid S_1 \rightarrow \alpha_1 \in P_1\} \cup \{S_3 \rightarrow \alpha_2 \mid S_2 \rightarrow \alpha_2 \in P_2\}$

G_3 – right linear language

and

$L(G_3) = L(G_1) \cup L(G_2)$

PROOF!!! Homework

ii. $G_4 = (N_4, \boldsymbol{\Sigma}, P_4, S_4)$

$N_4 = N_1 \cup N_2$; $S_4 = S_1$

$P_4 =$ $\{A \to aB \mid \text{if } A \to aB \in P_1\} \cup$
$\{A \to aS_2 \mid \text{if } A \to a \in P_1\} \cup P_2$

G_4 – right linear language
and
$L(G_4) = L(G_1)\, L(G_2)$

PROOF!!! Homework

iii. $G_5 = (N_5, \boldsymbol{\Sigma}, P_5, S_5)$

//IDEA: concatenate L_1 with itself

$N_4 = N_1 \cup \{S_5\}$;

$P_5 = \quad P_1 \cup \{S_5 \rightarrow \boldsymbol{\varepsilon}\} \cup$

$\{S_5 \rightarrow \boldsymbol{\alpha}_1 \mid S_1 \rightarrow \boldsymbol{\alpha}_1 \in P_1\} \cup$

$\{A \rightarrow aS_1 \mid \text{if } A \rightarrow a \in P_1\}$

G_5 – right linear language

and

$L(G_5) = L(G_1)^*$

PROOF!!! Homework

Theorem: *A language is a regular set if and only if is a right linear language*

Proof:

=> Apply lemma 1 and lemma 2

<= construct a system of regular exp equations where:

- Indeterminants – nonterminals
- Coefficients – terminals
- Equation for A: all the possible rewritings of A

Example: G=({S,A,B},{0,1}, P, S)

P: $S \to 0A \mid 1B \mid \varepsilon$

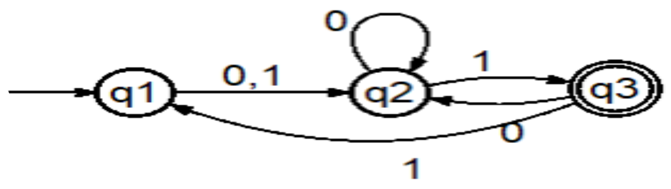
$A \to 0B \mid 1A$

$B \to 0S \mid 1$

$$\begin{cases} S = 0A + 1B + \boldsymbol{\varepsilon} \\ A = 0B + 1A \\ B = 0S + 1 \end{cases}$$

Regular exp = solution corresponding to S

Theorem: *A language is a regular set if and only if is accepted by a FA*



Proof:

=> Apply lemma 1 and lemma 2 (to follow, similar to RG)

<= construct a system of regular exp equations where:

- Indeterminants – states
- Coefficients – terminals
- Equation for A: all the possibilities that put the FA in state A
- Equation of the form: X=Xa+b => solution X=ba*

$$\begin{cases} q_1 = q_3 0 + \boldsymbol{\varepsilon} \\ q_2 = q_1 0 + q_1 1 + q_2 0 + q_3 0 \\ q_3 = q_2 1 \end{cases}$$

Regular exp = union of solutions corresponding to final states

Lemma 1': $\boldsymbol{\Phi}$,{$\boldsymbol{\varepsilon}$}, {a},$\forall a \in \boldsymbol{\Sigma}$ are accepted by FA

Reg exp	FA
$\boldsymbol{\Phi}$	$M = (Q, \boldsymbol{\Sigma}, \delta, q_{0,}\ \boldsymbol{\Phi})$
$\boldsymbol{\varepsilon}$	$M = (Q, \boldsymbol{\Sigma}, \boldsymbol{\Phi}, q_{0,}\ \{q_0\})$
$a, \forall a \in \boldsymbol{\Sigma}$	$M = (\{q_0,q_1\}, \boldsymbol{\Sigma}, \{\delta(q_0,a) = q_1\}, q_{0,}\ \{q_1\})$

Lemma 2':If L_1 and L_2 are accepted by a FA then: $L_1 \cup L_2$, L_1L_2 and L_1^* are accepted by FA

Proof:

$M_1 = (Q_1, \Sigma, \delta_1, q_{01}, F_1)$ such that $L_1 = L(M_1)$

$M_2 = (Q_2, \Sigma, \delta_2, q_{02}, F_2)$ such that $L_2 = L(M_2)$

$M_3 = (Q_3, \Sigma, \delta_3, q_{03}, F_3)$

$Q_3 = Q_1 \cup Q_2 \cup \{q_{03}\}$

$F_3 = F_1 \cup F_2 \cup \{q_{03} \mid \text{if } q_{01} \in F_1 \text{ or } q_{02} \in F_2\}$

$\delta_3 = \delta_1 \cup \delta_2 \cup \{\delta_3(q_{03},a) = p \mid \exists \delta_1(q_{01},a) = p\} \cup \{\delta_3(q_{03},a) = p \mid \exists \delta_2(q_{02},a) = p\}$

$L(M_3) = L(M_1) \cup L(M_2)$

PROOF!!! Homework

$M_4 = (Q_4, \boldsymbol{\Sigma}, \delta_4, q_{04}, F_4)$

$Q_4 = Q_1 \cup Q_2$; $\quad q_{04} = q_{01}$

$F_3 = F_2 \cup \{q \in F_1 \mid \text{if } q_{02} \in F_2\}$

$$\delta_3(q,a) = \begin{cases} \delta_1(q,a), \text{ if } q \in Q_1\text{-}F_1 \\ \delta_1(q,a) \cup \delta_2(q_{02},a) \text{ if } q \in F_1 \\ \delta_2(q,a), \text{ if } q \in Q_2 \end{cases}$$

$L(M_3) = L(M_1)L(M_2)$

PROOF!!! Homework

$M_5 = (Q_5, \boldsymbol{\Sigma}, \delta_5, q_{05}, F_5)$

$Q_5 = Q_1$; $q_{05} = q_{01}$

$F_5 = F_1 \cup \{q_{01}\}$

$$\delta_5(q,a) = \begin{cases} \delta_1(q,a), \text{ if } q \in Q_1 - F_1 \\ \delta_1(q,a) \cup \delta_1(q_{01},a) \text{ if } q \in F_1 \end{cases}$$

$L(M_3) = L(M_1)^*$

PROOF!!! Homework

Pumping Lemma

- Not all languages are regular
- How to decide if a language is regular or not?

- Idea: pump symbols

Example: $L = \{0^n1^n \mid n >= 0\}$

Theorem: (Pumping lemma, Bar-Hillel)

Let L be a regular language. $\exists p \in N$, such that if $w \in L$ with $|w|>p$, then

$w = xyz$, where $0<|y|<=p$

and

$xy^iz \in L, \forall i \geq 0$

Proof

L regular => $\exists$ M = (Q,$\boldsymbol{\Sigma}$,$\boldsymbol{\delta}$, q_0, F) such that L= L(M)

Let |Q| = p

If w $\in$ L(M): $(q_0,w) \vdash^* (q_f\ , \boldsymbol{\varepsilon})$, $q_f \in F$

and

|w|>=p

process at least p+1 symbols

p states

$\Rightarrow \exists$ q_1 that appear in at least 2 configurations

$(q_0,xyz) \vdash^* (q1,yz) \vdash^+ (q1,z) \vdash^* (q_f\ , \boldsymbol{\varepsilon})$, $q_f \in F$ => 0<=|y|<=p

Proof (cont)

$(q_0,xy^iz) \overset{*}{\vdash} (q_1,y^iz)$

$\overset{*}{\vdash} (q_1,y^{i-1}z)$

$\overset{*}{\vdash} \ldots$

$\overset{*}{\vdash} (q_1,yz)$

$\overset{*}{\vdash} (q_1, z)$

$\overset{*}{\vdash}(q_f , \boldsymbol{\varepsilon}) , q_f \in F$

So, if $w=xyz \in L$ then $xy^iz \in L$, for all $i>0$

If $i=0$: $(q_0,xz) \overset{*}{\vdash} (q_1, z) \overset{*}{\vdash}(q_f , \boldsymbol{\varepsilon}) , q_f \in F$

Example: $L = \{0^n 1^n \mid n >= 0\}$

Suppose L is regular => $w = xyz = 0^n1^n$

Consider all possible decomposition =>

Case 1. $y = 0^k$

$$xyz = 0^{n-k}0^k1^n;\ xy^iz = 0^{n-k}0^{ik}1^n \notin L$$

Case 2. $y = 1^k$

$$xyz = 0^n1^k1^{n-k};\ xy^iz = 0^n1^{ik}1^{n-k} \notin L$$

Case 3. $y = 0^k\,1^l$

$$xyz = 0^{n-k}0^k1^l1^{n-l};\ xy^iz = 0^{n-k}(0^k1^l)^i1^{n-l} \notin L$$

Case 4. $y = 0^k\,1^K$

$$xyz = 0^{n-k}0^k1^k1^{n-k};\ xy^iz = 0^{n-k}0^k1^k0^k1^k\ldots1^{n-l} \notin L$$

=> L is not regular