Finite Automata and Regular Languages (part IV)

Prof. Dan A. Simovici

UMB

We discuss the behavior of the class of regular languages with respect to several operations previously defined.
We prove that this class is the smallest class of languages that

- contains all finite languages, and
- is closed with respect to union, product, and Kleene closure.

The class of regular languages is closed with respect to the complement operations as shown by the next theorem.

Theorem

Let A be an alphabet. If $L \subseteq A^*$ is a regular language, then its complement $\bar{L} = A^* - L$ is also a regular language.

Proof.

Suppose that $\mathcal{M} = (A, Q, \delta, q_0, F)$ is an automaton such that $L = L(\mathcal{M})$. Define the automaton $\mathcal{M}' = (A, Q, \delta, q_0, Q - F)$; thus, only the set of final states of $\mathcal{M}'$ differs from the corresponding set for $\mathcal{M}$. We have $x \in L(\mathcal{M}')$ if and only if $\delta^*(q_0, x) \in Q - F$, or, equivalently, if and only if $\delta^*(q_0, x) \notin F$. Therefore $L(\mathcal{M}') = A^* - L$, which proves that $A^* - L$ is regular. □

Transition systems are very convenient for proving closure properties of the class $\mathcal{R}$. To make use of these devices we use the following technical result.

Lemma

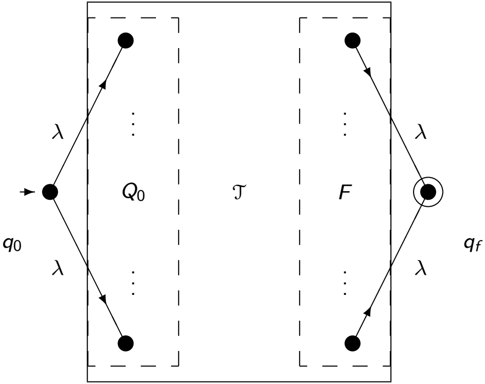
For every nonempty regular language $L \subseteq A^*$ there exists a transition system that accepts L and has a single initial state q_0 and a single final state q_f.

Proof

Since L is a nonempty regular language, there exists a transition system $\mathcal{T} = (A, Q, \theta, Q_0, F)$ such that $L(\mathcal{T}) = L$ and $F \neq \emptyset$. Define the transition system $\mathcal{T}' = (A, Q \cup \{q_0, q_f\}, \theta', \{q_0\}, \{q_f\})$, where $q_0, q_f \notin Q$. The relation θ' is given by:

$$\theta' = \theta \cup \{(q_0, \lambda, q) \mid q \in Q_0\} \cup \{(q, \lambda, q_f) \mid q \in F\}.$$

The graph of $\mathcal{T}'$ is represented below.



Clearly, $(q, x, q') \in \theta^*$ for some $q \in Q_0$ and $q' \in F$ if and only if $(q_0, \lambda x \lambda, q_f) = (q_0, x, q_f) \in \theta'^*$. Therefore, $L(\mathcal{T}) = L(\mathcal{T}')$.

Theorem

If L is a regular language, then L^R is also regular.

Proof.

Define a new transition system $\mathcal{T}' = (A, Q, \theta', \{q_f\}, \{q_0\})$, where

$$\theta' = \{(q_2, w, q_1) \mid (q_1, w, q_2) \in \theta\}.$$

We have $(q, x^R, q') \in \theta'^*$ if and only if $(q', x, q) \in \theta^*$. Therefore, $x \in L(\mathcal{T})$ if and only if $x^R \in L(\mathcal{T}')$, so $L(\mathcal{T}') = (L(\mathcal{T}))^R$. $\square$

Next, we show that the class of regular languages is closed under union.

Theorem

If L_0, L_1 are regular languages, then $L_0 \cup L_1$ is regular.

Proof.

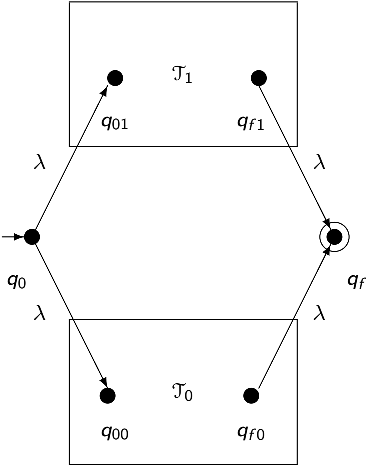
Without loss of generality, assume that both L_0 and L_1 are regular languages over the same alphabet A. Suppose that $L_i = L(\mathcal{T}_i)$, where $\mathcal{T}_i$, $i = 0, 1$, are two transition systems. We can assume that each $\mathcal{T}_i$ has a single initial and a single final state, $\mathcal{T}_i = (A, Q_i, \theta_i, \{q_{0i}\}, \{q_{fi}\})$ for $i = 0, 1$; also, assume that $Q_0 \cap Q_1 = \emptyset$.

□

Proof (cont'd)

Define a new transition system $\mathcal{T}' = (A, Q', \theta', \{q_0\}, \{q_f\})$ given by $Q' = Q_0 \cup Q_1 \cup \{q_0, q_f\}$, and

$$\theta = \theta_0 \cup \theta_1 \cup \{(q_0, \lambda, q_{00}), (q_0, \lambda, q_{01}), (q_{f0}, \lambda, q_f), (q_{f1}, \lambda, q_f)\}.$$



Proof (cont'd)

Since $Q_0 \cap Q_1 = \emptyset$, a path ϖ in $\mathcal{T}$ that joins q_0 to q_f exists in $\mathcal{T}$ if and only if that path passes through q_{00} and q_{f0}, or through q_{01} and q_{f1}. If x is the label of the path ϖ, then x belongs to $L(\mathcal{T}_0)$ or $L(\mathcal{T}_1)$, respectively. This amounts to $L(\mathcal{T}) = L(\mathcal{T}_0) \cup L(\mathcal{T}_1) = L_0 \cup L_1$, which implies that $L_0 \cup L_1$ is regular.

Corollary

The class of regular languages is closed under intersection. In other words, if L_0, L_1 are regular languages, then $L_0 \cap L_1$ is regular.

Proof.

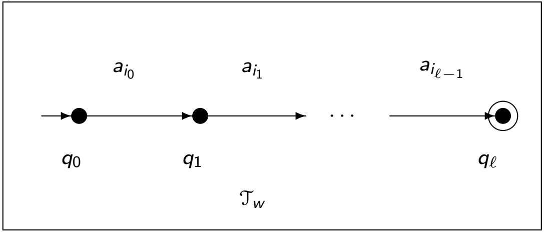
This statement follows immediately by previous theorems and by De Morgan's law. Specifically, $L_0 \cap L_1 = \overline{\overline{L_0} \cup \overline{L_1}}$, and each subexpression of the right hand side is regular if L_0 and L_1 are. □

Corollary

Every finite language over an alphabet A is regular.

Proof.

The empty language is clearly regular. Thus it suffices to show that one-word languages are regular. It is easy to see that if $L = \{w\}$, where $w = a_{i_0} \dots a_{i_{\ell-1}}$, then L is accepted by the transition system $\mathcal{T}_w$ given below



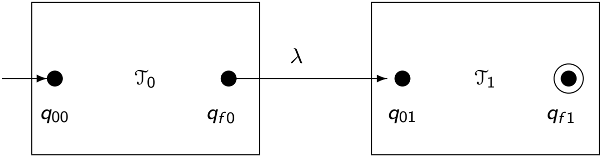
which implies the regularity of L. □

Theorem

If L_0, L_1 are regular languages, then $L_0 L_1$ is regular.

Proof.

Assume that both L_0 and L_1 are regular languages over the same alphabet A such that $L_i = L(\mathcal{T}_i)$, where $\mathcal{T}_i$, $i = 0, 1$, are two transition systems. We assume that each $\mathcal{T}_i$ has a single initial and a single final state, $\mathcal{T}_i = (A, Q_i, \theta_i, \{q_{0i}\}, \{q_{fi}\})$ for $i = 0, 1$; also, assume that $Q_0 \cap Q_1 = \emptyset$. Define the transition system $\mathcal{T} = (A, Q_0 \cup Q_1, \theta, \{q_{00}\}, \{q_{f1}\})$, where $\theta = \theta_0 \cup \theta_1 \cup \{(q_{f0}, \lambda, q_{01})\}$ □



Since $Q_0 \cap Q_1 = \emptyset$, to reach the state q_{f1} from the initial state q_{00} reading the symbols of the word x, the transition system $\mathcal{T}$ must pass through the states q_{f0} and q_{01} (via the null transition $(q_{f0}, \lambda, q_{01})$). This happens if and only if $x = uv$, where $(q_{00}, u, q_{f0}) \in \theta_0^*$ and $(q_{01}, v, q_{f1}) \in \theta_1^*$, so $L(\mathcal{T}) = L(\mathcal{T}_0)L(\mathcal{T}_1) = L_0L_1$. Hence, L_0L_1 is a regular language.

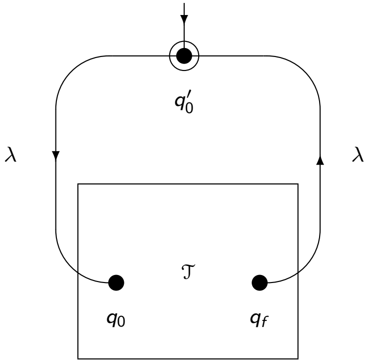
Theorem

If L is a regular language, then L^* is regular.

Proof.

Let $\mathcal{T} = (A, Q, \theta, \{q_0\}, \{q_f\})$ be a transition system such that $L = L(\mathcal{T})$. Define the transition system $\mathcal{T}' = (A, Q \cup \{q_0'\}, \theta', \{q_0'\}, \{q_0'\})$, where $\theta' = \theta \cup \{(q_f, \lambda, q_0'), (q_0', \lambda, q_0)\}$ and q_0' is a new state. □

Proof (cont'd)



Proof (cont'd)

We have $\lambda \in L(\mathcal{T}')$ because q_0' is both the initial and the final state of $\mathcal{T}'$. Further, if $w \in L(\mathcal{T})$, we have $(q_0, w, q_f) \in \theta^*$. Since both triples (q_0', λ, q_0) and (q_f, λ, q_0') belong to θ, we obtain $(q_0', w^k, q_0') \in \theta^*$ for every $k \in \mathbb{N}$, $k \geq 1$. Therefore, $L^k \subseteq L(\mathcal{T}')$ for $k \in \mathbb{N}$, so $L^* \subseteq L(\mathcal{T}')$.

Proof (cont'd)

Conversely, if $u \in L(\mathcal{T}')$, the transition system $\mathcal{T}'$ starts in q_0' and finishes in q_0' while reading the symbols of u. Let m be the number of times the transition system $\mathcal{T}'$ leaves the state q_0' while processing the word u. If $m = 0$, then $u = \lambda$. Otherwise, $m \geq 1$ and $\mathcal{T}'$ passes through the sequence of states: $q_0, \ldots, q_f, q_0', q_0, \ldots, q_f, q_0', \ldots, q_f$, where q_0' occurs m times. Here "passes through" means "enters and then leaves." This implies that we can write $u = u_0 \cdots u_{m+1}$, where $(q_0, u_i, q_f) \in \theta^*$ for $0 \leq i \leq m+1$. Thus, $w \in L^{m+1}$, so $L(\mathcal{T}') \subseteq L^*$. Hence, $L(\mathcal{T}') = L^*$.

Theorem

Let L be a regular language over the alphabet A. For every language K, both the right and the left quotients LK^{-1} and $K^{-1}L$ are regular.

Proof

We first deal with the left quotient. Let $\mathcal{T} = (A, Q, \theta, Q_0, F)$ be a transition system such that $L = L(\mathcal{T})$. Let $Q_K = \{q \in Q \mid (q_0, x, q) \in \theta^* \text{ for some } q_0 \in Q_0 \text{ and } x \in K\}$ and let $\mathcal{T}_K$ be the transition system $\mathcal{T}_K = (A, Q, \theta, Q_K, F)$. The following statements are easily seen to be equivalent:

1. $y \in K^{-1}L$;
2. $xy \in L$ for some $x \in K$;
3. $(q_0, xy, q') \in \theta^*$ for some $q_0 \in Q_0$ and $q' \in F$;
4. there is $q \in Q_K$ such that $(q_0, x, q) \in \theta^*$ and $(q, y, q') \in \theta^*$ for some $q_0 \in Q_0$ and $q' \in F$;
5. $y \in L(\mathcal{T}_K)$.

From these equivalences it follows that $K^{-1}L = L(\mathcal{T}_K)$, and thus $K^{-1}L$ is regular.

Proof (cont'd)

To make the argument for the right quotient, let P_K be the set $P_K = \{q \in Q \mid (q, z, q') \in \theta^* \text{ for some } q' \in F \text{ and } z \in K\}$. Define the transition system $\mathcal{T}^K$ as $\mathcal{T}^K = (A, Q, \theta, Q_0, P_K)$. We have the following equivalent statements:

1. $y \in L(\mathcal{T}^K)$;
2. $(q_0, y, q) \in \theta^*$ for some $q_0 \in Q_0$ and some $q \in P_K$;
3. $(q_0, y, q) \in \theta^*$ and $(q, z, q') \in \theta^*$ for some $q_0 \in Q_0$ and some $q' \in F$;
4. $(q_0, yz, q') \in \theta^*$ for some $q_0 \in Q_0$ and some $q' \in F$ and $z \in K$;
5. $yz \in L$ for some $z \in K$;
6. $y \in LK^{-1}$.

Therefore, $LK^{-1} = L(\mathcal{T}^K)$. This proves that the language LK^{-1} is regular. Note that this property of closure under quotients does not depend on the regularity of K.

Corollary

If $L \subseteq A^*$ is a regular language, then there exists a finite number of distinct left (right) quotients of the form $K^{-1}L$ (of the form LK^{-1}), where $K \subseteq A^*$.

Proof.

Suppose that $L = L(\mathcal{T})$, where $\mathcal{T} = (A, Q, \theta, Q_0, F)$. Note that if K, H are two languages such that $Q_K = Q_H$, then $K^{-1}L = H^{-1}L$. In other words, there are no more distinct left quotients than subsets of Q, which implies that the number of distinct left quotients of L is finite. □

Corollary

If $L \subseteq A^*$ is a regular language, then there exists a finite number of distinct left (right) derivatives of L.

Proof.

Follows immediately from the previous corollary by considering the quotients of L and singleton languages $K = \{x\}$ for $x \in A^*$. □

Corollary

If L is a regular language, then $\mathrm{PREF}(L)$, $\mathrm{SUFF}(L)$, *and* $\mathrm{INFIX}(L)$ *are all regular languages.*

Proof.

Follows from $\mathrm{SUFF}(L) = (A^*)^{-1}L$, $\mathrm{PREF}(L) = L\,(A^*)^{-1}$ and $\mathrm{INFIX}(L) = ((A^*)^{-1}\,L)\,(A^*)^{-1}$. □

Example

Using closure properties, it is easy to verify that if $\rho \subseteq A \times A$, then the language $L_\rho \subseteq A^*$ is regular. Indeed, we can write

$$A^* - L_\rho = \bigcup \{A^* a a' A^* \mid (a, a') \in (A \times A) - \rho\}.$$

Note that each language of the form $A^* a a' A^*$ is regular by. Furthermore, since A is a finite set, the right member of the equality is the union of a finite number of regular languages. Therefore, $A^* - L_\rho$ is regular, so implies that L_ρ is regular.