Programming Languages and Computation

Week 5: Regular Languages

** 1. Suppose $M = (Q, \Sigma, \delta, q_0, F)$ is a DFA. Construct a DFA $\overline{M}$ with $L(\overline{M}) = \{w \in \Sigma^* \mid w \notin L(M)\}$. You will not be able to answer this question with a diagram.

Solution

Let $\overline{F} = Q \setminus F$, i.e. $\{q \in Q \mid q \notin F\}$. Then define $\overline{M} = (Q, \Sigma, \delta, q_0, \overline{F})$.

** 2.

(a) Given a word $w \in \Sigma^*$, and a letter $a \in \Sigma$, let $\mathsf{erase}_a(w)$ be the word w but with any occurrences of a erased. For example, for alphabet $\Sigma = \{0,1,2\}$:

$$\begin{aligned}
\mathsf{erase}_1(01211210012) &= 022002 \\
\mathsf{erase}_1(101001222100) &= 00022200 \\
\mathsf{erase}_2(101001222100) &= 101001100 \\
\mathsf{erase}_2(101001100) &= 101001100
\end{aligned}$$

Define a function $\mathsf{erase}_a(R)$ that takes a regular expression R over Σ as input and returns a new regular expression S such that:

$$L(S) = \{\mathsf{erase}_a(w) \mid w \in L(R)\}$$

You need not justify your answer.

(b) Conclude that, if X is a regular language over Σ and $a \in \Sigma$, then $\{\mathsf{erase}_a(w) \mid w \in X\}$ is also a regular language over Σ.

Solution

We take R as input and then the regex that is output consists of replacing any occurrences of a by ϵ. We call this output $\mathsf{erase}_a(R)$ in the following definition:

$$\begin{aligned}
\mathsf{erase}_a(\emptyset) &= \emptyset \\
\mathsf{erase}_a(\epsilon) &= \epsilon \\
\mathsf{erase}_a(a) &= \epsilon \\
\mathsf{erase}_a(b) &= b \quad \text{if } a \neq b \\
\mathsf{erase}_a(R \cdot S) &= \mathsf{erase}_a(R) \cdot \mathsf{erase}_a(S) \\
\mathsf{erase}_a(R + S) &= \mathsf{erase}_a(R) + \mathsf{erase}_a(S) \\
\mathsf{erase}_a(R^*) &= (\mathsf{erase}_a(R))^*
\end{aligned}$$

(a) Suppose X is a regular language, then by definition it can be expressed as the language of a regular expression, say R. Then our transformation shows that $\{\mathsf{erase}_a(w) \mid w \in X\}$ can be expressed as a regular expression $\mathsf{erase}_a(R)$. Hence, by definition, it is regular.

** 3. Let rev(w) be the reverse of the word w, e.g. rev($abccd$) $= dccba$ and rev(ϵ) $= \epsilon$.

Show that language $S = \{w \in \{a,b\}^* \mid w = \text{rev}(w)\}$ is not regular.

Solution

Suppose S is regular. Then it follows from the pumping lemma that there is some length $p > 0$ such that, for any string s of length at least p, we can split it into three pieces the middle of which can be pumped. So let us consider $a^p b^p b^p a^p$. Since this word is clearly in S, it follows that $a^p b^p b^p a^p$ can be split as uvw with:

1. v not empty
2. $|uv| \leq p$
3. for all $i \in \mathbb{N}$: $uv^i w \in S$

By (2), we know that $uv = a^m$ for some $m \leq p$. By (3), we know that, for example $uvvw \in S$. Let's remember this fact as (*). Since v is not empty, it must be that $v = a^k$ for some $1 \leq k \leq m$. Hence, we have that $uvvw = a^{m-k}a^{2k}a^{p-m}b^p b^p a^p$, but $(m-k)+2k+(p-m) = p+k$ and $p+k > p$ (since $k \geq 1$). Hence, $uvvw \notin S$, contradicting (*) that we earlier deduced. Hence, we must withdraw our only supposition, which was that S is regular. Therefore, S is not regular.

** 4. Show that the language $\{a^n b^m \mid n = 2 * m\}$ over $\{a, b\}$ is not regular.

Solution

Let this language be A. Suppose A is regular. Then by the pumping lemma, there is some $p > 0$ such that, for all $s \in A$, s can be divided as uvw with:

1. $|uv| \leq p$, and
2. $|v| > 0$, and
3. $\forall i. uv^i w \in A$.

So take $s = a^{2p}b^p$ and consider any division satisfying 1–3. It must be, by (1), that $uv = a^k$ with $k \leq p$ and hence $v = a^\ell$ for (2) $0 < \ell \leq k \leq p$. Then by (3) with $i = 2$ we have $a^{2p+\ell}b^p \in A$, but $2p + \ell \neq 2p$ so $a^{2p+\ell}b^p \notin A$: contradiction.

*** 5. Prove that the language of squares (written in unary), $\{1^{n^2} \mid n \in \mathbb{N}\}$, is not regular.

Solution

Suppose S *were* regular, then it follows from the pumping lemma that there is a certain length of strings from S, say $p > 0$, at and beyond which we can guarantee repetition. So, let's consider 1^{p^2}, which is indeed a string of S with length at least p. The pumping lemma gives us that for this string (and others like it) the string can be split into three pieces: $1^{p^2} = uvw$. We don't know the exact split, but we are guaranteed that:

1. v is non-empty
2. $|uv| \leq p$
3. for all $i \in \mathbb{N}$: $uv^i w \in S$.

By (3), we know that, for example, $uvvw \in S$. But v is a nonempty string of length at most p. This means the word $uvvw = 1^{(p^2+|v|)}$ and $|v| \leq p$ (recall that $|v|$ is the length of v). However, $p^2 + |v|$ is not square, since it sits strictly between p^2 and $(p+1)^2$ (using $p \geq 1$ and $1 \leq |v| \leq p$):

$$p^2 < p^2 + |v| \leq p^2 + p < p^2 + 2p + 1 = (p+1)^2$$

Hence, we must withdraw our only assumption, namely that S is regular.

*** 6. Show that the following language over $\{0,1\}$ is regular and justify your answer:

$$\{0^k u 0^k \mid k \geq 1 \wedge u \in \{0,1\}^*\}$$

Solution

Call this language A. We claim that $A = L(00^*(0+1)^*00^*)$, and therefore it is regular.

Clearly, $A \subseteq L(00^*(0+1)^*00^*)$. Now will show that $L(00^*(0+1)^*00^*) \subseteq A$ (from both of these together it follows that, therefore, $A = L(00^*(0+1)^*00^*)$). So consider any word $w \in L(00^*(0+1)^*00^*)$. It follows that w has shape $0^i u 0^j$ for $i, j \geq 1$ and $u \in \{0,1\}^*$. Then let $k = \mathsf{min}(i,j)$ and it follows immediately that w has shape $0^k v 0^k$ for some $v \in \{0,1\}^*$. So, also $00^*(0+1)^*00^* \subseteq A$.