Assignment 1 solutions

Assume that $\Sigma = \{0, 1\}$ for every question on this assignment.

1. Prove that there are countably many finite languages over Σ.

Solution. There are several ways to answer this problem. One is to observe that every finite language is regular, and that there are countably many regular languages (as we observed in class). As every subset of a countable set is countable, the claim follows. A more direct solution to the problem is to define an onto function $f : \mathbb{N} \to \mathcal{F}$ for $\mathcal{F}$ denoting the set of all finite languages over Σ.

Here is one way (out of many) to do this. For each string $x \in \Sigma^*$ let us write $i(x)$ to denote the *index* (or position) of x in the list of all strings over Σ ordered lexicographically and starting with 1. For instance, $i(\varepsilon) = 1$, $i(0) = 2$, $i(1) = 3$, $i(00) = 4$, and so on. Also, for each $j \geq 1$ define p_j to be the j-th prime number: $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, and so on. Now define f as follows:

$$f(n) = \left\{ x \in \Sigma^* \, : \, p_{i(x)} \text{ divides } n + 1 \right\}.$$

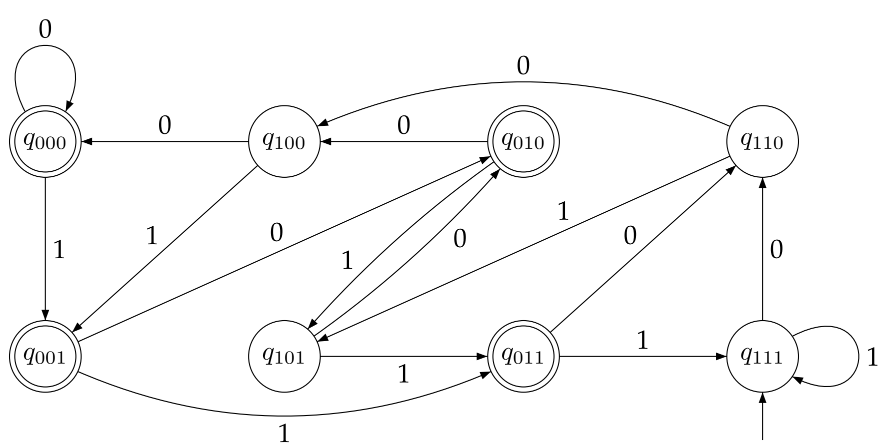
It is clear that each set $f(n)$ is finite because every positive integer has finitely many prime divisors. To see that f is onto, note that $f(0) = \varnothing$, and that for any nonempty finite set $S = \{x_1, \ldots, x_k\}$ we have $f(n) = S$ for $n = p_{i(x_1)} \cdots p_{i(x_k)} - 1$.

2. For each positive integer n define the language A_n over Σ as follows:

$$A_n = \{x \in \Sigma^* \, : \, |x| \geq n \text{ and the } n\text{-th symbol from the end of } x \text{ is } 0\}.$$

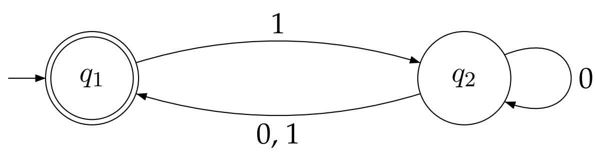
Give a DFA for A_3 that has the minimum number of states required for such a DFA.

Solution. The minimal number of states is 8, and (as is always true) the minimal DFA is unique:



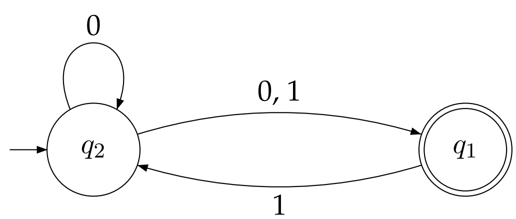
This diagram is similar to the diagram on page 51 in the course text [M. Sipser. *Introduction to the Theory of Computing (second edition)*. Thomson Course Technology, 2006], except that the accept and reject states are swapped and q_{111} is the start state. (It was not my intention to assign a problem whose solution is in the text, but at least it allows me to illustrate how you should cite sources in assignment solutions when it is appropriate.)

3. Consider the following NFA N:

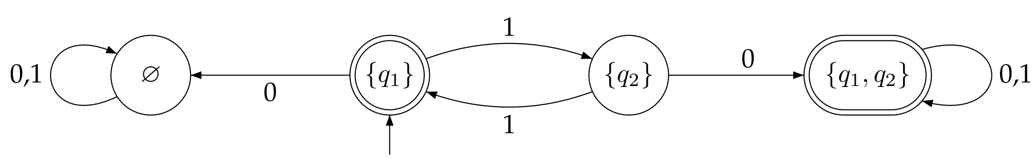


Give an NFA with just two states that recognizes the language $\overline{L(N)}$ over Σ.

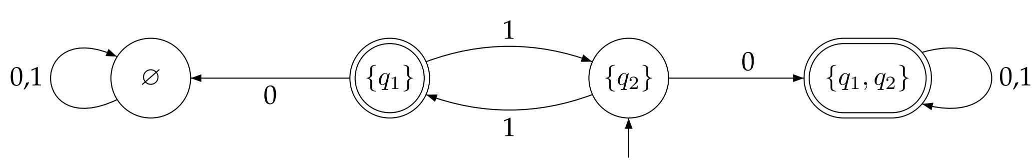
Solution. There is only one solution: the NFA is the same as N, except that q_2 is the start state rather than q_1:



One simple way to convince yourself that this NFA works is to convert the two NFAs to DFAs. When we convert N to a DFA we get this one:



The new NFA above, on the other hand, gives this DFA:



(The only change is the designation of the start state.) It is clear by inspection of these DFAs that they accept complementary languages.

4. Decide whether the following statements are true or false, and defend your answer: give a short proof for each statement you believe is true, and give a counter-example to each statement you believe is false.

   a. Let $A \subseteq \Sigma^*$. If A^* is regular then A is also regular.
   b. Let $A \subseteq \Sigma^*$ and $B \subseteq \Sigma^*$ be non-regular languages. Then AB is also non-regular.
   c. For every choice of languages A, B, and C over Σ it holds that $A(B \cap C) = AB \cap AC$.

d. Let $A \subseteq \Sigma^*$ be a regular language. Then there exists a DFA M with an even number of states such that $A = L(M)$.

Solution. Statement (a) is **false**. Consider, for instance, the language $A = \{x \in \Sigma^* : x = x^R\}$, which we proved is nonregular in the lecture. It holds that $A^* = \Sigma^*$, which is regular.

Statement (b) is **false**. For a counter-example we may let $C \subseteq \Sigma^*$ be any nonregular language and let

$$A = C \cup \{\varepsilon\} \quad \text{and} \quad B = \overline{C} \cup \{\varepsilon\}.$$

Then A and B are both nonregular, but $AB = \Sigma^*$, which is regular.

Statement (c) is **false**. One counter-example to the statement is obtained by letting $A = \Sigma^*$, $B = \{\varepsilon\}$, and $C = \{0\}$. Then $A(B \cap C) = \varnothing$, while $AB \cap AC = L((0 \cup 1)^* 0)$, which of course is not equal to the empty set.

Statement (d) is **true**. If A is regular, then by definition there exists a DFA D for A. If D has an even number of states, then we may take $M = D$. If D has an odd number of states, then we may take M to be a DFA formed by adding a new state to D, with the transitions out of this new state defined arbitrarily (e.g., both inducing a self-loop on this new state).

5. Let us say that a string x is obtained from a string w by *deletions* if it is possible to remove zero or more symbols from w so that just the string x remains.

   Suppose $A \subseteq \Sigma^*$ is a given regular language, and define

   $$B = \{x \in \Sigma^* : x \text{ is obtained from some string } w \in A \text{ by deletions}\}.$$

   Prove that B is regular.

Solution. Given that A is regular, there must exist a DFA $M = (Q, \Sigma, \delta, q_0, F)$ with $L(M) = A$. Define an NFA $N = (Q, \Sigma, \mu, q_0, F)$, with its transition function μ defined as

1. $\mu(q, \sigma) = \{\delta(q, \sigma)\}$ for each $q \in Q$ and $\sigma \in \Sigma$, and
2. $\mu(q, \varepsilon) = \{\delta(q, \sigma) : \sigma \in \Sigma\}$ for each $q \in Q$.

In other words, to obtain the state diagram of N we take the diagram for M and then include an additional ε-transition for each transition of M. It remains to prove that $L(N) = B$, as it follows from this fact, along with a theorem we proved in Lecture 3, that B is regular.

Arguably the claim that $L(N) = B$ is sufficiently clear that a proof could safely be omitted—I would not deduct points from a solution that simply stated that this equality is clear. Here is a high-level sketch of how such a proof might look.

First let us prove that $B \subseteq L(N)$. Consider any string $x \in B$, which is obtained from some string $w \in A = L(M)$ by deletions. We must show that $x \in L(N)$, and we may reason that this is so by referring to the sequence of states that witnesses the fact that w is accepted by M.

Next we must prove that $L(N) \subseteq B$, so we consider any string x that is accepted by N. By referring to the definition of acceptance for an NFA, and using the definition of N, we may obtain a string $w \in A$ from which x is obtained by deletions (where the deletions occur in positions corresponding to the ε-transitions). Thus, $x \in B$, and therefore $L(N) \subseteq B$ as required.