Technische Universität München **Winter term 2009/10**
I7
Prof. J. Esparza / J. Křetínský / M. Luttenberger **11. Februar 2010**

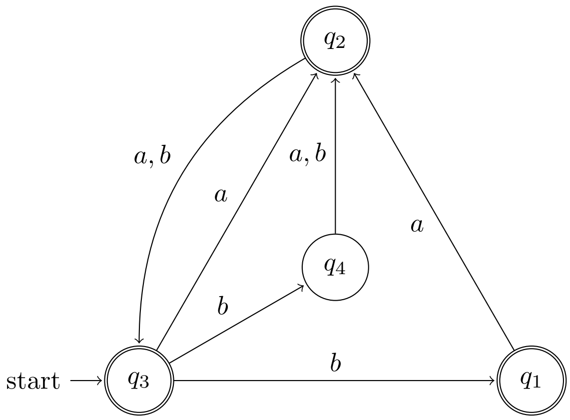
Solution

Automata and Formal Languages – Homework 5

Due 26.11.2009.

Exercise 5.1

Check whether the NFA depicted below recognizes Σ^* by means of the algorithm "UnivNFA" presented in the lecture.



Exercise 5.2

We define the following languages over the alphabet $\Sigma = \{a, b\}$:

- L_1 is the set of all words where between any two occurrences of b's there is at least one a.
- L_2 is the set of all words where every maximal sequence of consecutive a's has odd length.
- L_3 is the set of all words where a occurs only at even positions.
- L_4 is the set of all words where a occurs only at odd positions.
- L_5 is the set of all words of odd length.
- L_6 is the set of all words with an even number of a's.

Remark: For this exercise we assume that the first letter of a nonempty word is at position 1, e.g., $a \in L_4$, $a \notin L_3$.

Your task is to construct an FA, i.e., DFA or NFA or NFA-ε, for

$$L := (L_1 \setminus L_2) \cup \overline{(L_3 \triangle L_4) \cap L_5 \cap L_6} \text{ where } \triangle \text{ denotes the symmetric difference.}$$

while sticking to the following rules:

- You have to start from FAs for $L_1, \ldots, L_6$.
- Any further automaton has to be constructed from already existing automata via an algorithm introduced in the lecture, e.g., Bisim, BinOp, Comp, NFAtoDFA, etc.
- You are free to transform the propositional formula underlying the definition of L as you see fit.

Try to find an order on the construction steps which yields an FA for L with as few as possible states.

Exercise 2.2

For $n \in \mathbb{N}_0$ let msbf(n) be the language of all words over $\{0, 1\}$ which represent n w.r.t. to the *most significant bit first* representation where an arbitrary number of leading zeros is allowed. For example:

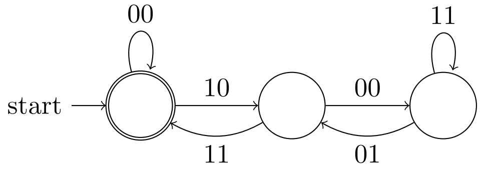
$$\text{msbf}(3) = \mathcal{L}(0^*11) \text{ and } \text{msbf}(0) = \mathcal{L}(0^*).$$

Similarly, let lsbf(n) denote the language of all *least significant bit first* representations of n with an arbitrary number of following zeros, e.g.:

$$\text{lsbf}(6) = \mathcal{L}(0110^*) \text{ and } \text{lsbf}(0) = \mathcal{L}(0^*).$$

(a) Construct and compare the minimal DFAs representing all even natural numbers $n \in \mathbb{N}_0$ w.r.t. the unary encoding (i.e., $n \mapsto a^n$), the msbf encoding, and the lsbf encoding.

(b) Consider the following FA $\mathcal{A}$ over the alphabet $\{00, 01, 10, 11\}$:



W.r.t. the msbf encoding, we may interpret any word $w \in \{00, 01, 10, 11\}^*$ as a pair of natural numbers $(X(w), Y(w)) \in \mathbb{N}_0 \times \mathbb{N}_0$. *Example*: (Underlined letters correspond to $Y(w)$.)

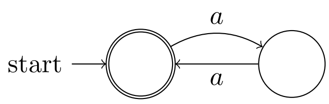
$$w = (00)^k001011 \to (0\underline{0})^k0\underline{0}1\underline{0}1\underline{1} \to (0^k011, 0^k001) \to (3, 1) = (X(w), Y(w))$$

- Find constants $a, b \in \mathbb{Z}$ such that $aX(w) + bY(w) = 0$ for all $w \in \mathcal{L}(\mathcal{A})$.

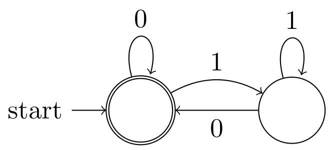
(c) Construct the minimal DFA representing the language $\{w \in \{0, 1\}^* \mid \text{msbf}^{-1}(w) \text{ is divisible by } 3\}$.

Solution:

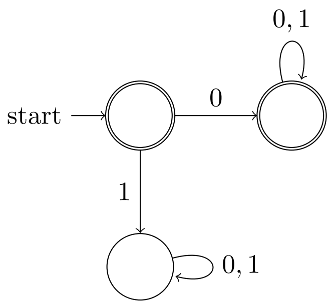
(a)
- Unary encoding:



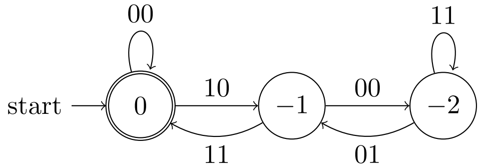
- MSBF encoding:



- LSBF encoding:



(b) We label the states as follows:



We show by induction on the length of a word $w \in \Sigma^*$ that for $\delta(0, w) = q$ we have $3Y(w) - X(w) = q$. Obviously, this will then imply that $3Y(w) - X(w) = 0$ for all $w \in \mathcal{L}(\mathcal{A})$.

Let l denote the length of a word w (w.r.t. the alphabet $\{00, 01, 10, 11\}$).

- $l = 0$:

  We have $w = \varepsilon$ and both $\delta(0, \varepsilon) = 0$ and $X(w) = 0 = Y(w)$.

- $l \to l + 1$:

  We may write w as uab with $a, b \in \{0, 1\}$. Note that we then have

  $$X(w) = 2X(u) + a \text{ and } Y(w) = 2Y(w) + b.$$

  Assume that $\delta(0, w) =: q$ is defined, otherwise there is nothing to show. Then also $\delta(0, u) =: q'$ is defined. By induction hypothesis we have $3Y(u) - X(u) = q'$.

  - Assume $q = 0 \wedge q' = 0$:

    Then $a = 0, b = 0$ has to hold, i.e., $X(w) = 2X(u)$ and $Y(w) = 2Y(u)$. Hence,

    $$3Y(w) - X(w) = 3 \cdot 2Y(u) - 2X(u) = 2(3Y(u) - X(u)) = 2 \cdot q' = 2 \cdot 0 = 0 = q.$$

  - Assume $q = 0 \wedge q' = -1$.

    Then $a = 1, b = 1$, and $X(w) = 2X(u) + 1$ and $Y(w) = 2Y(u) + 1$ subsequently follow, leading to:

    $$3Y(w) - X(w) = 3 \cdot (2Y(u) + 1) - (2X(u) + 1) = 2(3Y(u) - X(u)) + 2 = 2 \cdot q' + 2 = 2 \cdot (-1) + 2 = 0 = q.$$

  - Similarly, the remaining cases follow.

  One can even show that $\mathcal{A}$ accepts exactly those words w with $3Y(w) - X(w) = 0$. For this, let $w = uab$ be a word such that $\delta(0, u)$ is defined, but $\delta(0, w)$ is undefined, i.e., is the rejecting state.

  Assume $\delta(0, u) = 0$. Then either $ab = 01$ or $ab = 11$.

  - We only consider the case $ab = 01$, as the case $ab = 11$ is quite similar. Then by our preceding result we have $3Y(u) - X(u) = 0$, which leads to

    $$3Y(w) - X(w) = 3(2Y(u) + 1) - 2X(u) = 3.$$

    Assume we add a letter xy to w. In order to calculate $3Y(wxy) - X(wxy)$ we double the value $3Y(w) - X(w)$ and then add a number from $\{-1, 0, 2, 3\}$. Hence, $5 = 2(3Y(w) - X(w)) - 1 \leq 3Y(wxy) - X(wxy) \leq 2(3Y(w) - X(w)) + 3 = 9$, i.e., no matter how we extend w by some word $u \in \{00, 01, 10, 11\}^*$, we never will be able to obtain $3Y(wu) - X(wu)$.

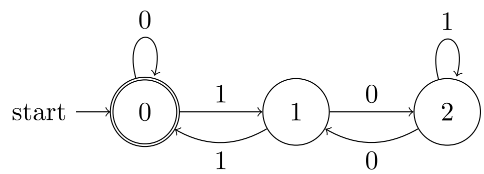
  Similarly, one shows for the other cases of $\delta(0, u)$ that all words leading to the rejecting state cannot satisfy the linear equation $3y - x = 0$.

(c) The idea is that the state reached after reading the word u corresponds to the remainder of the number represented by u when dividing by 3.

We therefore take as states $\{0, 1, 2\}$ with 0 the initial state and define

$$\delta(q, a) = 2q + a \pmod 3.$$

This yields the automaton:



Obviously, this automaton has to be minimal as two different states encode two different remainder classes. Note that we also obtain this automaton from the one of (b) by forgetting (projecting) the y-component.