Tree Languages and Applications

Sample solutions for the Exam, January 12, 2023

1 Residuals

(a) Each leaf guesses whether it is on a path of length n (and goes to q_1), or on some other path. Formally, our NFTA is $\langle\{q, q_1, \ldots, q_n\}, \mathcal{F}, \{q_n\}, \Delta\rangle$, where Δ contains, for all $i = 1, \ldots, n-1$:

$$a \to q \qquad a \to q_1 \qquad f(q,q) \to q \qquad f(q_i, q) \to q_{i+1} \qquad f(q, q_i) \to q_{i+1}$$

Alternative solution : It is possible to push the non-determinism into the f-transitions instead, with an automaton $\langle\{q_1, \ldots, q_n, q_{n+1}\}, \mathcal{F}, \{q_n\}, \Delta\rangle$ and, for all $i = 1, \ldots, n$ and $j = 1, \ldots, n+1$,

$$a \to q_1 \qquad f(q_i, q_j) \to q_{i+1} \qquad f(q_j, q_i) \to q_{i+1} \qquad f(q_{n+1}, q_{n+1}) \to q_{n+1}$$

The last transition here is important to prevent the automaton from blocking when all branches of some sub-tree are longer than n.

(b) The result certainly holds for $n = 1$ since one needs at least one state to accept any tree. For the rest, let $n > 1$.
Let $K := \{2, \ldots, n\}$. For each $I \subseteq K$ we can construct a tree t_I such that, for any $i \in K$, t_I has a branch of length i iff $i \in I$. Indeed, fix $t_I := t_I^{(2)}$, and

$$t_I^{(i)} = \begin{cases} f(a, t_I^{(i+1)}) & \text{if } i \in I \\ f(t_I^{(i+1)}, t_I^{(i+1)}) & \text{if } i \in K \setminus I \\ f(a,a) & \text{otherwise} \end{cases}$$

Suppose that we have a DFTA $\mathcal{A}$ accepting L_n with fewer than 2^{n-1} states. Then there must exist two different sets $I, J \subseteq K$ and a state q of $\mathcal{A}$ such that $t_I \to_{\mathcal{A}}^* q$ and $t_J \to_{\mathcal{A}}^* q$. Let i be the maximal index in the symmetric difference of I and J, and w.l.o.g. suppose that $i \in I \setminus J$.
We now consider the family of contexts $C_0 = x_1$ and $C_{k+1} = f(t_\emptyset, C_k)$, for $k \geq 0$. Then $C_{n-i}[t_I] \in L_n$ but $C_{n-i}[t_J] \notin L_n$. However, they are either both accepted or both rejected by $\mathcal{A}$, a contradiction.

(c) If L is recognizable, let $\mathcal{A} = \langle Q, \mathcal{F}, G, \Delta\rangle$ be a DFTA with n states accepting L, and let $L_q = \{\, t \in T(\mathcal{F}) \mid t \to_{\mathcal{A}}^* q \,\}$. For $C \in \mathcal{C}(\mathcal{F})$, let $Q_C = \{\, q \in Q \mid \exists q' \in G : C[q] \to_{\mathcal{A}}^* q' \,\}$. Then $C^{-1}L = \bigcup_{q \in Q_C} L_q$. Since $C^{-1}L$ is entirely determined by Q_C, we have $|R(L)| \leq 2^{|Q|}$.

(d) For $n \geq 2$, let $C \in \mathcal{C}(\mathcal{F})$ be a context whose variable position is p :
— If C contains a branch of length n (other than to p), then $C^{-1}L_n = T(\mathcal{F})$.
— If C contains no such branch and $0 \leq |p| < n$, then $C^{-1}L_n = L_{n-|p|}$.
— Otherwise, $C^{-1}L_n = \emptyset$.
Clearly, these $n+2$ residuals are all pairwise different. If $n = 1$, then the first case is not applicable, and there are only two residuals – this was a bug in the question, my apologies. (Unless one considers $C = a$ as a context, which stretches the definition.)

2 Prime decompositions

(a) We can construct a top-down automaton that mimicks the usual addition with carry, with just some extra checks to ensure that the right format is respected. For $i, j, k \in \{0, 1\}$, let us denote $[ijk]_s = (i + j + k) \bmod 2$ (the sum modulo 2) and $[ijk]_c = \lfloor (i + j + k)/2 \rfloor$ (the carry). Our T-NFTA is $\langle \{q, 0, 1\}, \mathcal{F}, \{q\}, \Delta \rangle$, with the following transitions, for all $i, j, k \in \{0, 1\}$:
— $q(j, k, [0, j, k]_s) \to [0, j, k]_c$;
— $i(j, k, [i, j, k]_s) \to [i, j, k]_c$;
— $i(\perp, k, [i, 0, k]_s) \to [i, 0, k]_c$ and $i(j, \perp, [i, j, 0]_s) \to [i, j, 0]_c$;
— $1(\perp, \perp, 1) \to 0$ and $0(\perp, \perp, \perp) \to \varepsilon$.
The accepting run on $\langle \widetilde{6}, \widetilde{3}, \widetilde{9} \rangle$ is as follows (parentheses omitted for readability) :

$$\begin{array}{ll} & \mathbf{q}\langle 011\rangle\langle 110\rangle\langle 1\perp 0\rangle\langle \perp\perp 1\rangle\langle \perp\perp\perp\rangle \\ \to & \langle 011\rangle\mathbf{0}\langle 110\rangle\langle 1\perp 0\rangle\langle \perp\perp 1\rangle\langle \perp\perp\perp\rangle \\ \to & \langle 011\rangle\langle 110\rangle\mathbf{1}\langle 1\perp 0\rangle\langle \perp\perp 1\rangle\langle \perp\perp\perp\rangle \\ \to & \langle 011\rangle\langle 110\rangle\langle 1\perp 0\rangle\mathbf{1}\langle \perp\perp 1\rangle\langle \perp\perp\perp\rangle \\ \to & \langle 011\rangle\langle 110\rangle\langle 1\perp 0\rangle\langle \perp\perp 1\rangle\mathbf{0}\langle \perp\perp\perp\rangle \\ \to & \langle 011\rangle\langle 110\rangle\langle 1\perp 0\rangle\langle \perp\perp 1\rangle\langle \perp\perp\perp\rangle \end{array}$$

(b) The following bottom-up automaton will do : $\langle \{q_\perp, q_0, q_1, q_f, q_+, q_*\}, \mathcal{F}, \{q_0, q_1, q_+\}, \Delta \rangle$, and Δ contains :

$$\perp \to q_\perp \qquad 0(q_\perp) \to q_0 \qquad 1(q_\perp) \to q_1 \qquad q_1 \to q_* \qquad 0(q_*) \to q_* \qquad 1(q_*) \to q_*$$

$$f(q_*, q_\perp) \to q_f \qquad f(q_*, q_f) \to q_f \qquad f(q_0, q_f) \to q_f \qquad 1(q_f) \to q_+$$

Here, q_0, q_1 recognize the encodings $\overline{0}, \overline{1}$, respectively, and q_+ recognizes the encodings of all other integers. In the prime factors, q_* recognizes $\widetilde{n}$, for any $n \geq 1$. The rules for f ensure that the highest prime factor has a non-zero multiple.

(c) In the following, for $a \in \mathcal{F}$, let $\Pi(a) = \{\langle a, \perp\rangle, \langle \perp, a\rangle, \langle a, a\rangle\}$.
When reading a tree encoding $\langle \overline{n}, \overline{m}, \overline{k} \rangle$, the main idea is of course to check that the prime multiples of k are the sums of those of n and m. The rest follows from the results in (a) and (b). There are some tedious technicalities to take care of :
— From (a), one can assume that there exists an DFTA that that reduces $\langle \widetilde{0}, \widetilde{0}, \widetilde{0} \rangle$ to q_0, and any other pair $\langle \widetilde{n'}, \widetilde{m'}, \widetilde{n' + m'} \rangle$ to q_*.
— Also, it is trivial to modify this DFTA to additionally treat an all-$\perp$ representation of either summand (but not both) as zero (in case where the maximal prime factor of n and m is different).
— We add the following transitions, for $\langle f_1, f_2 \rangle \in \Pi(f)$, where q_+ is accepting :

$$\langle \perp\perp\perp \rangle \to q_\perp \qquad \langle f_1 f_2 f \rangle(q_*, q_\perp) \to q_f \qquad \langle f_1 f_2 f \rangle(q_*, q_f) \to q_f$$

$$\langle f_1 f_2 f \rangle(q_0, q_f) \to q_f \qquad \langle 111 \rangle(q_f) \to q_+ \qquad \langle 010 \rangle(q_f) \to q_+ \qquad \langle 100 \rangle(q_f) \to q_+$$

— Also, one needs to handle the cases where n and m are both at most 1. This requires to recognize a finite number of additional trees, so can clearly be handled by an NFTA.
— Finally, the resulting automaton ought to be intersected with one verifying that all three projections are valid representations of some natural number, using (b).

3 Closures

(a) L_1 is not recognizable. To see this, consider the language L_3 of trees where a occurs to the left of the root and b to its right. L_3 is recognized by $\langle \{q_a, q_b, q_f\}, \mathcal{F}, \{q_f\}, \Delta \rangle$ with

$$a \to q_a \qquad f(q_a, q_a) \to q_a \qquad b \to q_b \qquad f(q_b, q_b) \to q_b \qquad f(q_a, q_b) \to q_f$$

If L_1 was recognizable, then so would $L_4 := L_1 \cap L_3$ be. But L_4 contains

$$f(\underbrace{f(f(\cdots(f(}_{i} a, a), \ldots), a), a), \underbrace{f(f(\cdots(f(}_{i} b, b), \ldots), b), b))$$

for every $i \geq 0$. It is now trivial to apply the pumping lemma to show that L_4 is not recognizable. But then neither is L_1.

(b) L_2 is recognizable by an NFTA $\langle \{q_a, q_b, q_f, q_r\}, \mathcal{F}, \{q_f\}, \Delta \rangle$ with the following rules :

$$a \to q_a \qquad b \to q_b \qquad f(q_a, q_b) \to q_f \qquad f(q_a, q_f) \to q_r \qquad f(q_r, q_b) \to q_f$$

(c) If L is a recognizable word language, then it is recognized by a morphism ϕ from monoid $\langle M, \cdot \rangle$. The tree language in question is recognized by the DFTA $\langle M, \mathcal{F}, \phi(L), \Delta \rangle$, where Δ contains $a \to \phi(a)$, $b \to \phi(b)$, and $f(m, m') \to m \cdot m'$ for all $m, m' \in M$.

(d) No. Let L_3 and L_4 as in the proof of (a). Again, L_2 is recognizable, but the associative closure of L_2 intersected with L_3 is L_4, which is not recognizable.

(e) No. L_2 is recognizable. The commutative and associative closure of L_2 is L_1, which is not recognizable.

(f) Yes. Let $\mathcal{A} = \langle Q, \mathcal{F}, G, \Delta \rangle$ an NFTA recognizing L. Then $\mathcal{A}' = \langle Q, \mathcal{F}, G, \Delta \cup \Delta' \rangle$ recognizes its commutative closure with $\Delta' = \{\, f(q, q') \to q'' \mid f(q', q) \to q'' \in \Delta \,\}$.