Automates d'arbre

TD n°1 : Recognizable Tree Languages and Finite Tree Automata*

Exercise 1 :

Let $\mathcal{F} = \{f(2), g(1), a(0)\}$. Give a DFTA and a top-down DFTA for the set $G(t)$ of ground instances of the term $t = f\big(f(a,x), g(y)\big)$ which is defined by :

$$G(t) = \Big\{ f\big(f(a,u), g(v)\big) \mid u, v \in T(\mathcal{F}) \Big\}$$

Solution:

- top-down DFTA : $Q = \{q_{f,1}, q_{f,2}, q_g, q_a, q_\top\}$, $I = \{q_{f,1}\}$ and $\Delta =$
  - ⋆ $q_{f,1}(f(x,y)) \longrightarrow f(q_{f,2}(x), q_g(y))$
  - ⋆ $q_{f_2}(f(x,y)) \longrightarrow f(q_a(x), q_\top(y))$
  - ⋆ $q_g(g(x)) \longrightarrow g(q_\top(x))$
  - ⋆ $q_a(a) \longrightarrow a$
  - ⋆ $q_\top(f(x,y)) \longrightarrow f(q_\top(x), q_\top(y))$
  - ⋆ $q_\top(g(x)) \longrightarrow g(q_\top(x))$
  - ⋆ $q_\top(a) \longrightarrow a$
- DFTA : $Q = \{q_a, q_f, q_g, q_\top, q_\perp\}$, $F = \{q_\top\}$ and $\Delta =$
  - ⋆ $a \longrightarrow q_a$
  - ⋆ $f(q_a, q) \longrightarrow q_f$ for all $q \in Q$
  - ⋆ $g(q) \longrightarrow q_g$ for all $q \in Q$
  - ⋆ $f(q_f, q_g) \longrightarrow q_\top$
  - ⋆ $f(q, q') \longrightarrow q_\perp$ for all $(q, q') \neq (q_a, \_), (q_f, q_g)$

Exercise 2 :

Are the following tree languages recognizable (by a bottom-up FTA) ?

- $\mathcal{F} = \{f(2), a(0)\}$ and $L = \{f(t,t) \mid t \in T(\mathcal{F})\}$.
- $\mathcal{F} = \{g(1), a(0)\}$ and L the set of ground terms of even height.
- $\mathcal{F} = \{f(2), g(1), a(0)\}$ and L the set of ground terms of even height.

Solution:

- No. By the pumping lemma.
- Yes.
- No. Remark that the pumping lemma does not apply ! Assume that it is recognizable by a NFTA with n states. Define :

  $$t_n = f(g^{2n+1}(a), g^{2n+2}(a))$$

  It has height $2n+2$ and so belongs to this language. So there exists an accepting run ρ for t_n. By the pigeonhole principle, there exists $k < k'$ such that $r(1.1^k) = r(1.1^{k'})$ and from that we deduce that for all $p \in \mathbb{N}$, the tree

  $$t_{n,p} = f(g^{2n+1+p(k'-k)}(a), g^{2n+2}(a))$$

  also has an accepting run. But $t_{n,2}$ has height $2(n + k' - k) + 1$ which is odd. Contradiction.

*taken from *Tree Automata Techniques and Applications*

Exercise 3 :

1) Let $\mathcal{E}$ be a finite set of linear terms on $T(\mathcal{F}, \mathcal{X})$. Prove that $Red(\mathcal{E}) = \{C[t\sigma] \mid C \in \mathcal{C}(\mathcal{F}), t \in \mathcal{E}, \sigma \text{ ground substitution}\}$ is recognizable.
2) Prove that if $\mathcal{E}$ contains only ground terms, then one can construct a DFTA recognizing $Red(\mathcal{E})$ whose number of states is at most $n + 2$, where n is the number of nodes of $\mathcal{E}$.

Solution:

1) Do the case where $\mathcal{E}$ is a singleton $\{t\}$, t linear (the general case can be deduced by finite union). $Red(\{t\}$ is recognized by the following NFTA : $Q = \{q_\perp\} \cup Pos(t)$, $F = \{\epsilon\}$ and $\Delta =$
   - ⋆ $f(q_1, ..., q_n) \longrightarrow q_\perp$ for all $f \in \mathcal{F}$, q_1, ..., $q_n \in Q$
   - ⋆ $q_\perp \longrightarrow p$ for all $p \in Pos(t)$ such that $t(p)$ is a variable
   - ⋆ $f(p.1, ..., p.n) \longrightarrow p$ if $t(p) = f$
   - ⋆ $f(q_1, ..., q_n) \longrightarrow \epsilon$ for all $f \in \mathcal{F}$ and q_1, ..., $q_n \in Q$ such that there exists $i \in \{1, ..., n\}$ such that $q_i = \epsilon$
2) Let $St(\mathcal{E})$ be the set of all subterms of a term in $\mathcal{E}$. Then the following DFTA works : $Q = \{q_t \mid t \in St(\mathcal{E})\} \cup \{q_\perp, q_\top\}$, $F = \{q_\top\}$ and $\Delta =$
   - ⋆ $f(q_{t_1}, ..., q_{t_n}) \longrightarrow q_{f(t_1,...,t_n)}$ if $f(t_1, ...t_n) \in St(\mathcal{E}) \setminus \mathcal{E}$
   - ⋆ $f(q_{t_1}, ..., q_{t_n}) \longrightarrow q_\top$ if $f(t_1, ..., t_n) \in \mathcal{E}$
   - ⋆ $f(q_{t_1}, ..., q_{t_n}) \longrightarrow q_\perp$ else
   - ⋆ $f(q_1, ..., q_n) \longrightarrow q_\top$ if there is at least one $q_i = q_\top$
   - ⋆ $f(q_1, ..., q_n) \longrightarrow q_\perp$ else

Exercise 4 :

Let $\mathcal{F} = \{f(2), a(0), b(0)\}$.

1) Let L_1 be the smallest set such that :
   - $f(a, b) \in L_1$
   - $t \in L_1 \Rightarrow f(a, f(t, b)) \in L_1$

   Prove that L_1 is recognizable.
2) Prove that $L_2 = \{t \in T(\mathcal{F}) \mid |t|_a = |t|_b\}$ is not recognizable.
3) Let L be recognizable on $\mathcal{F}$ and $C(L)$ be the closure of L by the congruence generated by the equation $f(x, y) = f(y, x)$. Prove that $C(L)$ is recognizable.
4) Let L be recognizable on $\mathcal{F}$ and $AC(L)$ be the closure of L by the congruence generated by the equations $f(x, y) = f(y, x)$ and $f(x, f(y, z)) = f(f(x, y), z)$. Prove that $AC(L)$ is not recognizable in general.
5) Let L be recognizable on $\mathcal{F}$ and $A(L)$ be the closure of L by the congruence generated by the equation $f(x, f(y, z)) = f(f(x, y), z)$. Prove that $A(L)$ is not recognizable in general.

Solution:

1) $Q = \{q_a, q_b, q_f, q_\top\}$, $F = \{q_\top\}$ and $\Delta =$
   - ⋆ $a \longrightarrow q_a$
   - ⋆ $b \longrightarrow q_b$
   - ⋆ $f(q_a, q_b) \longrightarrow q_\top$
   - ⋆ $f(q_\top, q_b) \longrightarrow q_f$
   - ⋆ $f(q_a, q_f) \longrightarrow q_\top$
2) By the pumping lemma.
3) Given a NFTA for L, construct a NFTA for $C(L)$ by adding for every rule of the form $f(q, q') \longrightarrow q''$, the rule $f(q', q) \longrightarrow q''$.
4) $AC(L_1) = L_2$.
5) $A(L_1)$ is not recognizable by the pumping lemma.

Exercise 5 :

Let $\mathcal{F} = \{f(2), g(1), a(0)\}$. Give a DFTA and a top-down NFTA for the set $M(t)$ of terms which have a ground instance of the term $t = f(a, g(x))$ as a subterm, ie. $M(t) = \{C[f(a, g(u))] \mid C \in \mathcal{C}(\mathcal{F}), u \in T(\mathcal{F})\}$.

Prove that $M(t)$ is not recognizable by a top-down DFTA or even a finite union of languages recognizable by a top-down DFTA.

hint : you can use without proof the following fact (prove it if you have time) : let t be a tree. The path language $\pi(t)$ is defined by :

- *if t is a constant, $\pi(t) = \{t\}$*
- *if $t = f(t_1, ..., t_n)$, $\pi(t) = \cup_{i=1}^{n}\{fiw \mid w \in \pi(t_i)\}$*

Let L be a tree language. The path language of L is $\pi(L) = \cup_{t\in L}\pi(t)$. The path closure of L is defined by $pathclosure(L) = \big\{t \mid \pi(t) \subseteq \pi(L)\big\}$. L is recognizable by a top-down DFTA iff L is recognizable and path closed, ie. $L = pathclosure(L)$.

Solution:

- top-down NFTA : $Q = \{q_0, q_\perp, q_a, q_g\}$, $I = \{q_0\}$ and $\Delta =$
  - ⋆ $q_0(f(x,y)) \longrightarrow f(q_\perp(x), q_0(y))$
  - ⋆ $q_0(f(x,y)) \longrightarrow f(q_0(x), q_\perp(y))$
  - ⋆ $q_\perp(f(x,y)) \longrightarrow f(q_\perp(x), q_\perp(y))$
  - ⋆ $q_\perp(g(x)) \longrightarrow g(q_\perp(x))$
  - ⋆ $q_\perp(a) \longrightarrow a$
  - ⋆ $q_0(g(x)) \longrightarrow g(q_0(x))$
  - ⋆ $q_0(f(x,y)) \longrightarrow f(q_a(x), q_g(y))$
  - ⋆ $q_a(a) \longrightarrow a$
  - ⋆ $q_g(g(x)) \longrightarrow g(q_\perp(x))$
- DFTA : $Q = \{q_a, q_g, q_\top, q_\perp\}$, $F = \{q_\top\}$ and $\Delta =$
  - ⋆ $a \longrightarrow q_a$
  - ⋆ $g(q_\top) \longrightarrow q_\top$
  - ⋆ $g(q) \longrightarrow q_g$ with $q \neq q_\top$
  - ⋆ $f(q, q') \longrightarrow q_\top$ if $(q, q') = (q_a, q_g)$ or $q = q_\top$ or $q' = q_\top$
  - ⋆ $f(q, q') \longrightarrow q_\perp$ else
- non-recognizable by a top-down DFTA : it is not path-closed because if it were the case, as $f(f(a, g(a)), a)$ and $f(a, f(a, g(a)))$ are in $M(t)$, then $f(a, a)$ would be in $M(t)$ too which is absurd.
- not a finite union of languages recognizable by a top-down DFTA : assume that $M(t)$ is a finite union $\bigcup_{k=1}^{n} L_k$ where L_k is recognizable by a top-down DFTA and in particular path-closed. Let t^p be defined by induction :
  — $t^0 = f(a, t)$
  — $t^{p+1} = f(t^p, a)$
  They all belong to $M(t)$ then at least one of the L_k contains two different t^p and t^q with, say, $p < q$. As L_k is path-closed, it must also contains s^p, where s^r is defined by :
  — $s^0 = f(a, a)$
  — $s^{r+1} = f(s^r, a)$
  which does not belong to $M(t)$ (because, for example, it contains only fs and as). Contradiction.