Automates d’arbre

TD n°4 : Unranked trees, encodings

Exercise 1 :
Let $\Sigma = \{a, b, c\}$. Define a NFHA $A$ such that $L(A)$ is the set of all trees such that there exist two nodes labelled by $b$ whose greatest common ancestor is labelled by $c$.

**Solution:**
$Q = \{q_{\bot}, q_{\top}, q_b\}$, $F = \{q_{\top}\}$ and $\Delta =$

- $b(q_{\bot}) \rightarrow q_b$
- $\_ (q_{\bot}, q_b, q_{\bot}) \rightarrow q_b$
- $c(q_{\bot}, q_b, q_{\bot}) \rightarrow q_{\top}$
- $\_ (q_{\bot}, q_{\top}, q_{\bot}) \rightarrow q_{\top}$
- $\_ (q_{\bot}) \rightarrow q_{\bot}$

Exercise 2 :
Let $\Sigma = \{a, b\}$. Define a DFHA $A$ such that $L(A)$ is the set of all trees such that "for every leaf labeled with $a$, there is an ancestor from which there is a path whose nodes are labeled with $b$". Here "ancestor" means strict ancestor and "from which there is a path" means that there is a path from a son of this ancestor to a leaf.

**Solution:**
$Q = \{q_a, q_b, q_{\top}\}$, $F = \{q_b, q_{\top}\}$ and $\Delta =$

- $a(\epsilon) \rightarrow q_a$
- $a(q_{\top}) \rightarrow q_{\top}$
- $a((q_{\top} + q_a)^* q_a (q_{\top} + q_a)^*) \rightarrow q_a$
- $b(q_{\top}) \rightarrow q_{\top}$
- $b(\epsilon) \rightarrow q_b$
- $a(Q^* q_b Q^*) \rightarrow q_{\top}$
- $b(Q^* q_b Q^*) \rightarrow q_b$
- $b((q_{\top} + q_a)^* q_a (q_{\top} + q_a)^*) \rightarrow q_a$

**Definition** (extension encoding).
Let $t$ be an unranked tree on $\Sigma$. Let $F_{ext}^\Sigma = \{ @ (2) \} \cup \{ a(0) \mid a \in \Sigma \}$. We define the ranked tree $\text{ext}(t)$ by induction on the size of $t$ by :

- for $a \in \Sigma$, $\text{ext}(a) = a$
- if $t = a(t_1, ..., t_n)$ with $n \geq 1$, $\text{ext}(t) = @ (\text{ext}(a(t_1, ..., t_{n-1})), \text{ext}(t_n))$

that is $\text{ext}(a(t_1, ..., t_n))$ is equal to :
Exercise 3:
Give the extension encoding of:

```
    a
   /|
  /  |
 b  c  d
```

Solution:

```
    a
   /|
  /  |
 b  c  d
```

Exercise 4:
Let $L$ be a language of unranked trees. Prove that $L$ is recognizable by a NFHA iff $\text{ext}(L)$ is recognizable by a NFTA.

Solution:

\[
\mathcal{A}' = (Q', F_{\text{ext}}', \Delta', F')
\]

where:
- $Q' = \bigcup_{(a,q)} P_{a,q}$
- $F' = \bigcup_{(a,q)|q \in F} F_{a,q}$
- $\Delta' =
  \begin{align*}
  & * a \rightarrow p_0^{a,q} \text{ for all } (a,q) \\
  & \otimes (p, p') \rightarrow p'' \text{ if } p, p'' \in P_{b,q}, p' \in F_{a,q'} \text{ with } \delta_{b,q}(p, q') = p'' \text{ for some } b, q, a, q'
  \end{align*}

\(\Rightarrow\) Let $\mathcal{A} = (Q, \Sigma, \Delta, F)$ be a NFHA recognizing $L$ such that there is exactly one rule of the form $a(L_{a,q}) \rightarrow q$ for all $(a, q)$ and let $B_{a,q} = (P_{a,q}, Q, p_0^{a,q}, \delta_{a,q}, F_{a,q})$ a deterministic automaton recognizing $L_{a,q}$. We construct the expected NFTA this way:

\[
\mathcal{A}' = (Q', F_{\text{ext}}', \Delta', F')
\]

\(\Leftarrow\) Let $\mathcal{A} = (Q, F_{\text{ext}}', F, \Delta)$ be a NFTA recognizing $\text{ext}(L)$. We construct the expected NFHA this way:

\[
\mathcal{A}' = (Q, \Sigma, F, \Delta')
\]
where for all \((a, q), a(R_{a,q}) \rightarrow q \in \Delta'\) where \(R_{a,q}\) is the language recognized by the automaton:

\[
B_{a,q} = \langle Q, Q, I_{a,q}, F_{a,q}, \Delta_{a,q} \rangle
\]

with:

- \(I_{a,q} = \{p \in Q \mid a \rightarrow p \in \Delta\}\)
- \(F_{a,q} = \{q\} \) if \(q \in F\) or if there exists \(q', q''\) such that \((q', q) \rightarrow q'' \in \Delta\)
- \(\emptyset\) else
- \(\Delta_{a,q} = \{(q_1, q_2, q_3) \mid \oplus(q_1, q_2) \rightarrow q_3 \in \Delta\}\)