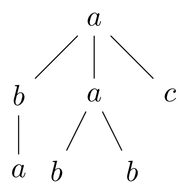
Automates d'arbre

TD n°5 : PDL & Document Type Definitions (DTD)

Exercise 1 :

Let t be the tree :



Which formulae are satisfied by t ?

1. $\phi_1 = \neg a \vee \langle\downarrow\rangle \Big(\neg \langle\leftarrow\rangle \top \wedge b \wedge \langle\rightarrow^*\rangle (c \wedge \neg \langle\rightarrow\rangle \top)\Big)$
2. $\phi_2 = \neg a \vee \langle\downarrow\rangle \Big(\neg \langle\leftarrow\rangle \top \wedge b \wedge \langle(\rightarrow; c?)^*\rangle (\neg \langle\rightarrow\rangle \top)\Big)$
3. $\phi_3 = \langle(a?; \downarrow)^*\rangle (a \wedge \neg \langle\downarrow\rangle \top)$

Exercise 2 :

Fix an alphabet $\mathcal{F}$. Give a PDL path formula π satisfying both conditions :

- for every tree t and every position p of t, there exists exactly one position q of t such that $(p, q) \in [\![\pi]\!]_t$ (π defines a function on positions).
- for every tree t and every position p of t, $(p, q) \in [\![\pi^\star]\!]_t$ iff q is a position of t such that $t(q) = t(p)$.

Let us recall the MSO logic mentioned in the previous lecture. Formulae are generated by the following grammar :

$$\phi ::= x = y \mid x \in X \mid x \downarrow y \mid x \rightarrow y \mid P_a(x) \mid \neg\phi \mid \phi \vee \phi \mid \exists x.\phi \mid \exists X.\phi$$

where x, y are first order variables, X is a second order variable and a is an element of a fixed (unranked) alphabet. If ϕ is a MSO formula with free variables $x_1, \ldots x_n$, $X_1, \ldots, X_p$, t is an unranked tree, $w_1, \ldots, w_n$ are positions of t and $S_1, \ldots, S_p$ are sets of positions of t, we define $t, \{x_i \mapsto w_i, X_j \mapsto S_j\} \models \phi$ by induction on the size of ϕ :

— $t, \{x \mapsto w, y \mapsto w'\} \models x = y$ iff $w = w'$
— $t, \{x \mapsto w, X \mapsto S\} \models x \in X$ iff $w \in S$
— $t, \{x \mapsto w, y \mapsto w'\} \models x \downarrow y$ iff there is an integer i such that $w' = w.i$
— $t, \{x \mapsto w, y \mapsto w'\} \models x \rightarrow y$ iff there are a position p of t and an integer i such that $w = p.i$ and $w' = p.(i + 1)$
— $t, \{x \mapsto w\} \models P_a(x)$ iff $t(w) = a$
— $t, \{x_i \mapsto w_i, X_j \mapsto S_j\} \models \neg\phi$ iff $t, \{x_i \mapsto w_i, X_j \mapsto S_j\} \not\models \phi$
— $t, \{x_i \mapsto w_i, X_j \mapsto S_j\} \models \phi_1 \vee \phi_2$ iff $t, \{x_i \mapsto w_i, X_j \mapsto S_j\} \models \phi_1$ or $t, \{x_i \mapsto w_i, X_j \mapsto S_j\} \models \phi_2$
— $t, \{x_i \mapsto w_i, X_j \mapsto S_j\} \models \exists x.\phi$ iff there is a position w of t such that $t, \{x_i \mapsto w_i, x \mapsto w, X_j \mapsto S_j\} \models \phi$
— $t, \{x_i \mapsto w_i, X_j \mapsto S_j\} \models \exists X.\phi$ iff there is a set S of positions of t such that $t, \{x_i \mapsto w_i, X_j \mapsto S_j, X \mapsto S\} \models \phi$.

Exercise 3 :

Give a translation of PDL into MSO which preserves models. That is, given a position formula ϕ (resp. a path formula π), construct a MSO formula $\tilde{\phi}$ (resp. $\tilde{\pi}$) whose set of free variables is $\{x\}$ (resp. $\{x, y\}$) such that $t, w \models \phi$ iff $t, \{x \mapsto w\} \models \tilde{\phi}$ (resp. $t, w, w' \models \pi$ iff $t, \{x \mapsto w, y \mapsto w'\} \models \tilde{\pi}$).

Remark. One does not need the whole MSO power in this translation. In fact, if one restricts to conditional PDL, i.e., where iterations are restricted to those of the form $(\alpha; \phi?)^*$ where $\alpha \in \{\rightarrow, \downarrow, \uparrow, \leftarrow\}$, then one can translate it into $FO(\rightarrow^*, \downarrow^*, P_a)$.

Definition (DTD)**.** A *DTD* is a tuple (Σ, s, δ) with $s \in \Sigma$ (the start symbol) and $\delta : \Sigma \longrightarrow Rat_\Sigma$, where Rat_Σ is the set of rational expressions on Σ.

Definition (Language of a DTD)**.** The language of a DTD $D = (\Sigma, s, \delta)$ is the language of the NHTA $\mathcal{A} = (Q_D, \Sigma, Q_{D,f}, \Delta_D)$ where :
— the set of states $Q_D = \Sigma$
— the set of final states $Q_{D,f} = \{s\}$
— the transitions are the $a(\delta(a)) \longrightarrow a$ for $a \in \Sigma$ (the left 'a's are the letter a and the right one is the state a)

Observe that it is deterministic.

Exercise 4 :

Give a recognizable language whose trees have the same label on the root which is not the language of a DTD.

Exercise 5 :

1) Assume given a rational expression E. Produce a PDL path formula π_E such that for every tree t, and every w, w' positions of t, $t, w, w' \models \pi_E$ iff there is a position p of t and integers $i \leq j$ such that $w = p.i$, $w' = p.j$ and $t(p.(i+1)) \ldots t(p.j)$ belongs to the language of E.
2) Produce a PDL state formula ϕ_E such that for every tree t and position w of t, $t, w \models \phi_E$ iff there is a position p of t and an integer i such that $w = p.i$ and $t(p.0) \ldots t(p.i)$ belongs to the language of E.
3) Deduce that from a DTD, one can produce a PDL formula which defines the same language.

Definition (Marking, 1-unambiguous rational expression)**.** The *marking* $m(E)$ of a rational expression on Σ is the rational expression on $\Sigma \times \mathbb{N}$ obtained by replacing the ith occurrence of a in E by $(a, i-1)$ for all $a \in \Sigma$. For example :

$$m((a+b)^* b (ab)^*) = ((a,0) + (b,0))^* (b,1)((a,1)(b,2))^*$$

We say that E is *1-unambiguous* if for all words u, v and w in $\Sigma \times \mathbb{N}$, for all symbols a, b in Σ and all integers i, j, if $u(a,i)v$ and $u(b,j)w \in L(m(E))$ then $(a,i) \neq (b,j)$ implies $a \neq b$. We say that a DTD is *1-unambiguous* if for all $a \in \Sigma$, $\delta(a)$ is 1-unambiguous.

Remark. There are languages that are not definable by a 1-unambiguous rational expression but it is decidable in polynomial time wether a language is definable by such an expression. See A. Brüggemann-Klein, *Formal Models in Document Processing.*

Exercise 6 :

Let E be a regular expression on Σ. Define the following sets :

— $\text{first}(E) = \{a \in \Sigma \mid \exists w \in \Sigma^*,\ aw \in L(E)\}$

— $\text{last}(E) = \{a \in \Sigma \mid \exists w \in \Sigma^*,\ wa \in L(E)\}$

— $\text{follow}(E, a) = \{b \in \Sigma \mid \exists v,\ w \in \Sigma^*,\ vabw \in L(E)\}$

— $\text{sym}(E) = \{a \in \Sigma$ appearing in $E\}$

1) Prove that for all $n \geq 1$, $x_1 ... x_n \in L(m(E))$ iff the three following conditions hold :
   — $x_1 \in \text{first}(m(E))$
   — $x_n \in \text{last}(m(E))$
   — for all $1 \leq i < n$, $x_{i+1} \in \text{follow}(m(E), x_i)$
2) Prove that if E is 1-unambiguous then :
   — for all $(a, i), (b, j) \in \text{first}(m(E))$, $(a, i) \neq (b, j)$ implies $a \neq b$
   — for all $(a, i) \in \text{sym}(m(E))$, $(b, j), (c, k) \in \text{follow}(m(E), (a, i))$, $(b, j) \neq (c, k)$ implies $b \neq c$
3) Let E be 1-unambiguous. Construct in polynomial time on the size of E a deterministic automaton that recognizes $L(E)$.
4) Deduce that the inclusion of 1-unambiguous DTD is in P ?