Tree Automata and Applications: TD6
Nested Words

Emile Contal
http://econtal.perso.math.cnrs.fr/teaching

October 25, 2013

Definition 1. A *nested word* on the alphabet A is a word on the alphabet $A \cup A_c \cup A_r$ where $A_c = \{\langle a \mid a \in A\}$ and $A_r = \{a\rangle \mid a \in A\}$ are mutually disjoint and disjoints from A. If w is a nested word on A and i a position on w such that $w_i \in A_c$, we write $i \xrightarrow{w} j$ if j the smallest integer strictly greater that i (or ∞) such that $w_j \in A_r$ and $\big|\{k \mid w_k \in A_c, i < k < j\}\big| = \big|\{k \mid w_k \in A_r, i < k < j\}\big|$. Likewise, when $w_i \in A_r$, we write $j \xleftarrow{w} i$ if j is the greater integer (or $-\infty$) such that $w_j \in A_c$ and $\big|\{k \,|w|_k \in A_c, j < k < i\}\big| = \big|\{k \mid w_k \in A_r, j < k < i\}\big|$.

For example, $\langle aab\langle ba\rangle\langle aaa\rangle a\rangle$ is a nested word w such that $1 \xrightarrow{w} 9$, $4 \xrightarrow{w} 5$, $6 \xrightarrow{w} 8$. And $x = ab\rangle\langle a\langle aa\rangle a$ is such that $-\infty \xleftarrow{x} 2$, $3 \xrightarrow{x} +\infty$.

Definition 2. A *visibly pushdown automaton* is a pushdown automaton $\mathcal{A} = (\mathcal{Q}, \Sigma, \Gamma, \delta, \mathcal{Q}_0, \mathcal{Q}_f)$ where $\Sigma = A \cup A_c \cup A_r$ is the input alphabet and Γ is the stack alphabet, such that,

- when reading a letter from A_c, $\mathcal{A}$ pushes a letter on the stack and change its state,
- when reading a letter from A, $\mathcal{A}$ changes its state regardless of the stack,
- when reading a letter from A_r, $\mathcal{A}$ changes its state according to the top of the stack and pop off the top of the stack (or it does nothing if the stack is empty).

Formally, the transition relation δ verifies:

$$\delta \subseteq \big(Q \times A_c \times \Gamma \times Q\big) \cup \big(Q \times A \times Q\big) \cup \big(Q \times A_r \times (\Gamma \cup \{\bot\}) \times Q\big) \ .$$

A computation of $\mathcal{A}$ on $w = w_1 \cdots w_n$ is a sequence $(q_0, \gamma_0), \ldots, (q_n, \gamma_n)$ such that $\gamma_i \in \Gamma^* \perp$ and,

- if $w_i \in A$, then $(q_{i-1}, w_i, q_i) \in \delta$ and $\gamma_{i+1} = \gamma_i$,
- if $w_i \in A_c$, then $(q_i, w_i, c, q_{i+1}) \in \delta$ and $\gamma_{i+1} = c\gamma_i$,
- if $w_i \in A_r$, then $(q_i, w_i, c, q_{i+1}) \in \delta$ and $\gamma_i = c\gamma_{i+1}$ or $c = \perp$ and $\gamma_i = \gamma_{i+1}$.

A computation is accepting if $q_n \in \mathcal{Q}_f$. The accepting language is the set of words for which it exists an accepting computation.

Exercise 1. 1. We first consider the *well formed* nested words, that is the words w for which there is no index i such that $i \xrightarrow{w} +\infty$ or $-\infty \xleftarrow{w} i$.

   Show that for each well formed w we can build a tree t_w labelled by $A \cup \{\bullet\}$ such that the fonction $Y : T(A \cup \{\bullet\}) \to A^*$ defined by:

   - $Y\big(\bullet(t_1, \ldots, t_n)\big) = Y(t_1 \cdots t_n)$,
   - $Y(a \cdot t_w) = a \cdot Y(t_w)$ if a is a leaf,
   - $Y\big(a(u_1, \ldots, u_m, b) \cdot t_w\big) = \langle a Y(u_1 \cdots u_m) b \rangle Y(t_w)$ if b is a leaf and $a, b \in A$,

   verifies $Y(t_w) = w$ and for all well formed nested word language L accepted by a visibly pushdown automaton, the tree language $t_L = \{t_w \mid w \in L\}$ is recognizable.

2. Generalize the previous construction for arbitrary languages accepted by visibly pushdown automaton. Give a recognizable tree language E_0 such that:

   - For all language L accepted by a visibly pushdown automaton, then $t_L \subseteq E_0$ and t_L is recognizable.
   - If E is a recognizable tree language and $E \subseteq E_0$, then $Y(E)$ is a language accepted by a visibly pushdown automaton.

3. Show that the class of languages accepted by visibly pushdown automaton is closed by complement.