

ניהול נתונים באינטרנט – תרגיל 1

1. 3 שגיאות בקבצי ה-XML וה-DTD:
 - a. לא מוגדר שורש, כלומר לא מוגדר DOCTYPE (לא בקובץ ה-XML ולא ב-XML).
 - b. קובץ ה-XML לא מכיל פירוט האלמנט COST.
 - c. ב-PART שתגית ה-ITEM שלו היא Video Card, לא מופיע אלמנט ה-MODEL שהוגדר ב-XML.
2. פתרון:
 - a. בקובץ question2.dtd המצורף.
 - b. השאלות:
 - (i) //shop[price < 10 and not(@name = preceding :: shop[price < 10]/@name)]/@name
 - (ii) //shop[price > 3]/ancestor :: book/@id
 - (iii) //author[contains(text(), ' ')]/ancestor :: book/@id
 - (iv) //publish_date[matches(text(), '[2-9][0-9]{3}-.*')]/ancestor :: book/author/text()
3. תוצאות בקבצים המצורפים - 2bi.xml, עבור $i \in \{1, 2, 3, 4\}$:
 - a. ה-Document Order נקבע לפי ה-Pre-Order של העץ, כלומר:

$$A \rightarrow B \rightarrow E \rightarrow J \rightarrow K \rightarrow F \rightarrow L \rightarrow C \rightarrow G \rightarrow M \rightarrow D \rightarrow H \rightarrow N \rightarrow O \rightarrow I$$
 - b. במסמך המצורף, 3b.xml.
 - c. הרצות הצירים:
 - (i) //G/preceding – sibling :: * $\rightarrow \emptyset$
 - (ii) //C/descendant :: * $\rightarrow \{G, M\}$
 - (iii) //F/following :: * $\rightarrow \{C, G, M, D, H, N, O, I\}$
 - (iv) //N/ancestor – or – self :: * $\rightarrow \{A, D, H, N\}$
 - (v) //M/preceding :: * $\rightarrow \{B, E, J, K, F, L\}$
 - (vi) //K/parent :: */following :: * $\rightarrow \{F, L, C, G, M, D, H, N, O, I\}$
4. פתרון:
 - a. נגדיר אוטומט עץ Bottom-up, $T = (\Sigma, Q, F, \delta_0, \delta_1)$, באופן הבא:

$$\Sigma = \{G, B, R, U\}, s.t. G = GREEN, B = BLUE, R = RED, U = UNDEF$$

$$Q = \{q_0, q_G, q_B, q_R\}; F = \{q_G, q_B, q_R\}$$

$$\delta_0: \Sigma \rightarrow P(Q); \delta_1: \Sigma \times P(Q) \rightarrow P(Q)$$

denote $S = \{G, B, R\}$. Let A be a finite set, then denote: $q_A = \bigcup_{a \in A} \{q_a\}$

$$\delta_0(C) = \{q_C\}, \text{ for } C \in S$$

$$\forall C \in S: \delta_1(U, \{q_C\}) = q_S - \{q_C\}$$

$$\forall C_1, C_2 \in S; C_1 \neq C_2: \delta_1(U, \{q_{C_1}, q_{C_2}\}) = q_S - \{q_{C_1}, q_{C_2}\}$$

$$\delta_1(U, q_S = \{q_G, q_B, q_R\}) = \delta_1(U, \{q_0\}) = \{q_0\}$$

הסבר:

לכל צבע אפשרי קיים מצב, שכן האוטומט יקבל אמ"מ השורש שייך ל- $P(F)$, כלומר ניתן לצבוע אותו בצבע כלשהו. כמו-כן, נשים לב כי מהגדרת δ_1 , אם לצומת פנימי כלשהו, U , קיימים ילדים אשר צבועים בשלושת הצבעים, דהיינו נמצאים ב- q_S , לא ניתן לצבוע את U ונעבור למצב q_0 שאיננו מקבל (מצב זה הוא גם בור, שכן לא ניתן לצאת ממצב זה מהגדרת δ_1).

b. נגדיר אוטומט $T' = (\Sigma', Q', \{q_i\}, F', \delta, \delta_{leaf})$, $top - down$ כך שמתקיים:

$$\Sigma' = \{G, B, R, U\}; Q' = \{q_i, q_R, q_G, q_B, q_{rej}, q_{acc}\}, s.t. q_i \text{ is the initial state}$$

$$F' = \{q_{acc}\}; \text{Let us denote } Q_C = \{q_R, q_B, q_G\}, \text{ and } \mathcal{L}(Q') \text{ reg. lang. of states}$$

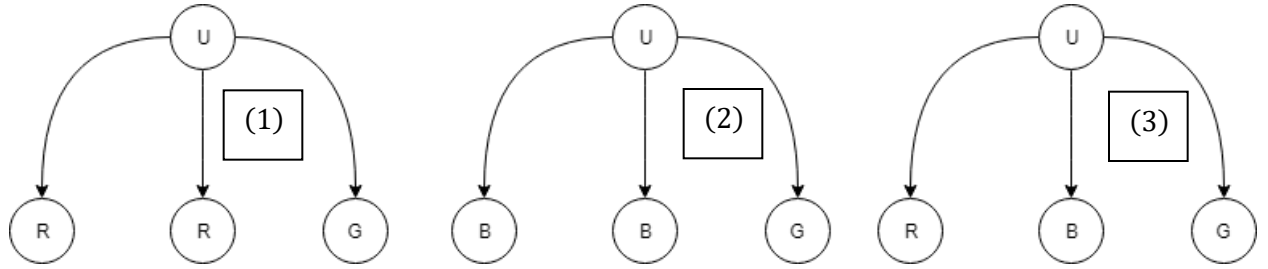
$$\delta: \Sigma' \times \mathcal{L}(Q') \rightarrow P(Q_C^3), s.t. Q_C^3 = Q_C \times Q_C \times Q_C; \delta_{leaf}: \Sigma' \times Q' \rightarrow Q'$$

$$\begin{aligned} \delta(U, q_i) &= Q_C^3 - \{(q_R, q_B, q_G), (q_B, q_R, q_G), (q_R, q_G, q_B), (q_G, q_B, q_R), (q_G, q_R, q_B), (q_B, q_G, q_R)\} \\ \delta(U, q_R) &= \{(q_G, q_B, q_G), (q_B, q_G, q_G), (q_G, q_G, q_B), (q_G, q_B, q_B), (q_G, q_B, q_B), (q_B, q_G, q_B), (q_G, q_G, q_G), (q_B, q_B, q_B)\} \\ \delta(U, q_B) &= \{(q_G, q_R, q_G), (q_R, q_G, q_G), (q_G, q_G, q_R), (q_G, q_R, q_R), (q_G, q_R, q_R), (q_R, q_G, q_R), (q_G, q_G, q_G), (q_R, q_R, q_R)\} \\ \delta(U, q_G) &= \{(q_R, q_B, q_R), (q_B, q_R, q_R), (q_R, q_R, q_B), (q_R, q_B, q_B), (q_R, q_B, q_B), (q_B, q_R, q_B), (q_R, q_R, q_R), (q_B, q_B, q_B)\} \\ \forall X \in \{R, G, B\}: \delta_{leaf}(X, q_X) &= q_{acc} \\ \forall X, Y \in \{R, G, B\}; Y \neq X: \delta_{leaf}(X, q_Y) &= q_{rej} \end{aligned}$$

הסבר:

- השורש, מהמצב ההתחלתי, יכול להגדיר כל מצב חוקי של צביעה, כלומר, לכל היותר 2 צבעים שונים לילדיו.
- כל צומת פנימי, יכול להגדיר צביעה ע"י כל מצב שהוא עצמו אינו נמצא בו, שרק במקרה זה נקבל מצב חוקי.
- העלים כבר "צבועים" אזי – יכנסו למצב מקבל אמ"מ המצב שלהם תואם לצבע שלהם.

c. לא ניתן לזהות ע"י אוטומט עץ $Top - down$ דטרמיניסטי. נניח בשלילה שניתן לזהות, כלומר קיים אוטומט עץ כני"ל, $T'' = (\{G, B, R, U\}, Q'', q_i, F'', \delta'', \delta''_{leaf})$, המקבל את שפת כל העצים שניתנים לצביעה. יהיו העצים:



יהי המצב ההתחלתי q_i . ונניח בה"ה כי:

$$\delta''(U, q_i) = (q_1, q_2, q_3), s.t. q_1, q_2, q_3 \in Q''$$

נשים לב כי עצים (1), (2) בשפה ועץ (3) אינו בשפה. נבחן את ריצת האוטומט.

ריצת האוטומט על עץ (1):

$$\delta''_{leaf}(R, q_1) = q_1^{acc}, \delta''_{leaf}(R, q_2) = q_2^{acc}, \delta''_{leaf}(G, q_3) = q_3^{acc}, s.t. \forall j \in \{1, 2, 3\}: q_j^{acc} \in F''$$

באופן דומה, עבור עץ (2):

$$\delta''_{leaf}(B, q_1) = q_4^{acc}, \delta''_{leaf}(B, q_2) = q_5^{acc}, \delta''_{leaf}(G, q_3) = q_3^{acc}, s.t. \forall j \in \{4, 5\}: q_j^{acc} \in F''$$

העלים במצבים מקבלים, שכן מהנחת השלילה T'' מקבל את השפה.

ריצת האוטומט על עץ (3):

$$\delta''_{leaf}(R, q_1) = q_1^{acc}, \delta''_{leaf}(B, q_2) = q_5^{acc}, \delta''_{leaf}(G, q_3) = q_3^{acc}$$

נשים לב כי כל העלים במצב מקבל, לכן האוטומט יקבל את עץ (3) בסתירה לכך שאינו בשפה. מכאן שלא קיים אוטומט עץ $Top - down$ דטרמיניסטי המקבל את השפה.