

אוטומטים ושפות פורמליות - תרגולים

מבוסס על תרגולים של מר גולדגביכט עומר, אוניברסיטת בר-אילן 2012.

שיעור 1

הגדרות:

א"ב: אוסף סופי ולא ריק של סימנים/אותיות/תווים. נסמן אותו באות Σ .

דוגמאות: $\{0, 1, \dots, 9\}$, $\{a, \dots, z\}$

מחרוזת/מילה: מילה הינה סדרה סופית של סימנים מא"ב כלשהו. למשל מעל $\{a, \dots, z\}$. נסמן אותה ב- w .

מילה ריקה: מילה זו היא סדרה של 0 תווים ונסמן אותה ב- ε .

אורך מילה: מספר התווים שיש במילה. נסמן את האורך של מילה ב- $|w|$.

שרשור מילים: יהיו w_1, w_2 מילים מעל א"ב Σ . אזי $w = w_1 w_2$ נקרא שרשור של w_1, w_2 .
למשל: $w_1 = ab$, $w_2 = b$ אזי $w = abb$.

העלאה של א"ב בחזקה: עבור Σ נתון נגדיר: Σ^n - קבוצת כל המילים מעל Σ באורך n .
למשל: $\Sigma^0 = \{\varepsilon\}$, $\Sigma^1 = \{0, 1\}$, $\Sigma^3 = \{000, 001, 010, 011, 100, 101, 110, 111\}$
 Σ^* :

קבוצת כל המילים מעל א"ב Σ . (נשים לב: $\Sigma^* = \bigcup_{n=0}^{\infty} \Sigma^n = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \dots$)

שפה: זוהי תת קבוצה של מילים מתוך Σ^* . (נשים לב כי גם $\emptyset$ וגם Σ^* הן שפות)

אות: תו ששייך לא"ב ונסמנו: $\sigma \in \Sigma$. מס' המופעים של σ ב- w הוא: $\#_{\sigma}(w)$

הערה: אם לא מצויין במפורש, ε לא חייב להיות שייך לשפה! כמו כן, $\emptyset \neq \{\varepsilon\}$

פעולות על שפות

פעולות שאפשר לעשות על קבוצות: חיתוך, איחוד, חיסור, משלים, הפרש סימטרי וכו'

בנוסף נגדיר שרשרות שפות כך: יהיו $L_1 \subseteq \Sigma_1^*$ ו- $L_2 \subseteq \Sigma_2^*$ שרשרות $L_1 \cdot L_2 \subseteq (\Sigma_1 \cup \Sigma_2)^*$ מוגדר באופן הבא: $L_1 \cdot L_2 = \{w = w_1 \cdot w_2 \mid w_1 \in L_1, w_2 \in L_2\}$

דוגמאות:

$$L_1 \cdot L_2 = \{00, 0a, 0b, 10, 1a, 1b\}, \quad L_1 = \{0, 1\}, \quad \Sigma_1 = \{0, 1\} \\ L_2 = \{0, a, b\}, \quad \Sigma_2 = \{0, a, b\} \quad (1)$$

$$(2) \quad L \cdot \emptyset = \emptyset \cdot L = \emptyset : L$$

$$(3) \quad L \cdot \varepsilon = \varepsilon \cdot L = L : L$$

הערה: נשים לב ששרשרות איננה פעולה קומטטיבית!

חזקות של שפות: לכל $L \subseteq \Sigma^*$, נגדיר באופן רקורסיבי:

$$L^n = \{w = w_1 \cdot w_2 \cdot \dots \cdot w_n \mid \forall 1 \leq i \leq n, w_i \in L\}, \quad \text{כלומר, } \begin{cases} L^0 = \{\varepsilon\} \\ L^1 = \{L\} \\ L^n = L^{n-1} \cdot L \end{cases}$$

$$\text{נגדיר: } * \text{-קליני: } L^* = \bigcup_{n=0}^{\infty} L^n$$

$$\text{דוגמאות: } (1) \quad L^* = \{0^i \mid i \geq 0\} \quad L = \{0\}$$

$$(2) \quad L = \{a, ab\} \quad \Sigma = \{a, b\} \quad \text{האם } w = aabaab \text{ שייכת ל-} L^*?$$

$$\text{כן: } w \in L^4 \subseteq \bigcup_{n=0}^{\infty} L^n = L^*$$

$$(3) \quad \emptyset^* = \{\varepsilon\}$$

$$\text{נגדיר } L^+ : L^+ = \bigcup_{n=1}^{\infty} L^n$$

$$\text{מתי } L^+ = L^*? \quad \text{אם } \varepsilon \in L \text{ אז } L^+ = L^*.$$

תרגיל

יהיו L_1, L_2, L_3 שפות מעל Σ הוכח או הפרך:

$$L_1 \cap L_2 L_3 = (L_1 \cap L_2) (L_1 \cap L_3) \quad (1)$$

הפרכה: $L_1 = \{a, b\}, L_2 = \{a\}, L_3 = \{b\}$

$$L_1 \cap L_2 L_3 = \{ab\}$$

$$\underbrace{(L_1 \cap L_2)}_{\emptyset} \underbrace{(L_1 \cap L_3)}_{\emptyset} = \emptyset$$

$$L_1 (L_2 \cup L_3) = L_1 L_2 \cup L_1 L_3 \quad (2)$$

הוכחה: תהי $w \in L_1 (L_2 \cup L_3) \iff w = yz : y, z \in \Sigma^*$ קיימים ומתקיים:
 $((z \in L_3 \wedge y \in L_1) \vee (z \in L_2 \wedge y \in L_1)) \iff ((z \in L_3 \vee z \in L_2) \wedge y \in L_1)$
 $\iff (w \in L_1 L_2 \cup L_1 L_3) \iff (w \in L_1 L_3 \vee w \in L_1 L_2) \iff$ כנדרש.

$$(L_1 \cup L_2) L_3 = L_1 L_3 \cap L_2 L_3 \quad (3)$$

הפרכה 1 (עם שפות סופיות): $\Sigma = \{a, b, c\}, L_1 = \{a, b, c\}, L_2 = \{a, c\}, L_3 = \{ba, a\}$

$$(L_1 \cap L_2) L_3 = \{cba, ca\}$$

$$L_1 L_3 \cap L_2 L_3 = \{abba, aba, cba, ca\} \cap \{aba, aa, cba, ca\} = \{aba, cba, ca\}$$

הפרכה 2: (עם שפות אינסופיות): $\Sigma = \{0\}, L_1 = \{0^n \mid n \text{ is even}\}, L_2 = \{0^n \mid n \text{ is odd}\}, L_3 = \Sigma^*$,

$$(L_1 \cap L_2) L_3 = \emptyset \cdot L_3 = \emptyset$$

$$L_1 L_3 \cap L_2 L_3 = \Sigma^* \cap L_2 L_3 \neq \emptyset$$

הסבר: מכיוון ש- L_2, L_3 אינן קבוצות ריקות, שרשור $L_2 L_3$ אינו קבוצה ריקה ולכן $\Sigma^* \cap L_2 L_3 \neq \emptyset$

הגדרה: נגדיר פונקציה $reverse : \Sigma^* \rightarrow \Sigma^*$ באופן הבא: $reverse(\sigma_1 \sigma_2 \dots \sigma_n) = \begin{cases} \varepsilon & n = 0 \\ \sigma_n \sigma_{n-1} \dots \sigma_1 & n > 0 \end{cases}$

נשמך בקיצור עבור מילה w , $w^R = reverse(w)$

תרגיל:

הוכח: לכל $x, y \in \Sigma^*$ מתקיים $(xy)^R = y^R x^R$

הוכחה: אם $x = \varepsilon$ אזי $x^R = \varepsilon$ ומתקיים $(xy)^R = (y)^R = y^R \varepsilon = y^R x^R$
 אם $y = \varepsilon$ אזי כמו מקודם.

אחרת $x, y \neq \varepsilon$ ונסמן $x = x_1 x_2 \dots x_n$, $y = y_1 y_2 \dots y_m$ ומתקיים:
 $(xy)^R = (x_1 x_2 \dots x_n y_1 y_2 \dots y_m)^R = y_m y_{m-1} \dots y_1 x_n x_{n-1} \dots x_1 = y^R x^R$ כנדרש.

תרגיל:

תנו דוגמה לקבוצה אינסופית A של שפות מעל הא"ב $\Sigma = \{0, 1\}$ כל שמתקיימים התנאים הבאים:

(1) כל חיתוך של מס' סופי של שפות אינו ריק.

(2) קיימת תת-קבוצה (אינסופית) של A כך שחיתוך השפות בה הוא ריק.

פתרון: נגדיר $L_k = \{0^n \mid n \geq k\}$ ו- $A = \bigcup_{k=0}^{\infty} \{L_k\} = \{L_0, L_1, L_2, \dots\}$

הוכחת תנאי 1: נתסכל בתת-קבוצה סופית של A , $A' \subseteq A$, נניח בה"כ ש- i_m הוא האינדקס הגדול ביותר. אזי חיתוך כל איברי A' הוא $L_{i_m} = \{0^n \mid n \geq i_m\}$ ואכן אינו ריק, כלומר $\bigcap_{L \in A'} L \neq \emptyset$

הוכחת תנאי 2: נסתכל בתת-קבוצה אינסופית כלשהי $A'' \subseteq A$. נניח בשלילה ש- $\bigcap_{L \in A''} L \neq \emptyset$ אזי קיימת

מילה 0^i ($i \geq 0$) ב- $\bigcap_{L \in A''} L$ מכיוון ש- A'' אינסופית אזי קיים $j > i$ כך ש- $L_j \in A''$

אבל לפי ההגדרה של L_j , $0^i \notin L_j$, בסתירה לכך ש- $0^i \in \bigcap_{L \in A''} L$

(למעשה הוכחנו משהו יותר חזק, שלכל תת-קבוצה אינסופית חיתוך השפות הוא ריק, ולא רק קיום תת-קבוצה כזו)

שיעור 2

תזכורת: $L^* = \bigcup_{n=0}^{\infty} L^n$

תרגיל:

הוכח: $(L^*)^* = L^*$

כיוון א': $(L^*)^* \subseteq L^*$

יהי $x \in (L^*)^*$ אזי קיים $n: x \in (L^*)^n$,

$\Leftarrow$ נשים לב כי $x = y_1 \dots y_n$ כך ש- $y_i \in L^*$ $\forall 1 \leq i \leq n$ לכל y_i קיים $m_i: y_i \in L^{m_i}$

$\Leftarrow$ $y_i = z_{i_1} z_{i_2} \dots z_{i_{m_i}}$ כך ש- $z_{i_j} \in L$

$$x \in L^{\sum_{i=1}^n m_i} \stackrel{=}{=} L^k \subseteq \bigcup_{n=0}^{\infty} L^n \Leftarrow x = \underbrace{z_{1_1} z_{1_2} \dots z_{1_{m_1}}}_{y_1} \underbrace{z_{2_1} z_{2_2} \dots z_{2_{m_2}}}_{y_2} \dots \underbrace{z_{n_1} z_{n_2} \dots z_{n_{m_n}}}_{y_n} \Leftarrow$$

$$x \in L^* \Leftarrow L^*$$

כיוון ב': $L^* \subseteq (L^*)^*$. יהי $x \in L^*$ $\Leftarrow x \in (L^*)^*$ $\Leftarrow x \in L^* = (L^*)^1 \subseteq \bigcup_{n=0}^{\infty} (L^*)^n = (L^*)^*$ כדרוש.

אוטומט סופי דטרמיניסטי

נקרא גם אס"ד, מוגדר ע"י החמישייה $\{\sum, Q, q_0, F, \delta\}$ כאשר: $\sum$ זהו הא"ב, Q רשימת מצבים, q_0 מצב התחלתי, $F \subseteq Q$ קבוצת מצבים מקבלים ו- δ פונקציית המעבר $\delta: Q \times \sum \rightarrow Q$ כך ש- $\delta(q, \sigma) = q'$ שם נמצאים במצב q וקוראים את האות σ , עוברים למצב q' .

$$\begin{aligned} \sum &= \{a, b\} \\ Q &= \{q_0, q_1\} \\ q_0 & \\ F &= \{q_0\} \end{aligned} \quad \text{דוגמא:}$$

$Q \setminus \sum$	a	b
q_0	q_1	q_1
q_1	q_0	q_0

את δ אפשר לייצג בטבלה באופן הבא:

האוטומט דטרמיניסטי, ולכן לכל מצב $q \in Q$ ולכל אות $\sigma \in \sum$ מוגדר מעבר $\delta(q, \sigma)$

הרחבה של δ למילים: נגדיר $\delta^*: Q \times \sum^* \rightarrow Q$ באופן רקורסיבי: $\delta^*(q, \varepsilon) = q$
 $\delta^*(q, w\sigma) = \delta(\delta^*(q, w), \sigma)$

שפת האוטומט: השפה שהאוטומט מקבל. מילה $w \in \sum^*$ מתקבלת ע"י אס"ד A אם $\delta^*(q_0, w) \in F$ לכן
השפה ש- A מקבל היא: $L(A) = \{w \mid \delta^*(q_0, w) \in F\}$. בדוגמה הקודמת $L(A) = \{w \mid |w| \bmod 2 = 0\}$

תרגילים:

בנה אס"ד מעל $\sum = \{a, b\}$ עבור השפות הבאות:

$$L_1 = \{w \mid |w| \bmod 2 = 1\} \quad (1) \quad F = \{q_1\}$$

$$L_2 = \{w \mid w \text{ contains the sequence 'ab'}\} \quad (2)$$

$$L_3 = \{wab \mid w \in \Sigma^* (\text{w ends with the sequence 'ab'})\} \quad (3)$$

$$L_4 = \{w \mid w \in \Sigma^* (\text{w doesn't ends with the sequence 'ab'})\} \quad (4)$$

$$. L_4 = \overline{L_3} = \Sigma^* - L_3$$

$$\delta_{L_3}(q, \sigma) = \delta_{L_4}(q, \sigma) : q \in Q, \sigma \in \Sigma \text{ ולכל } Q_{L_4} = Q_{L_3}, F_{L_4} = Q_{L_3} - F_{L_3} \text{ כך:}$$

$$L_5 = \{w\sigma \mid w \in \Sigma^*, \sigma \in \Sigma\} \quad (5)$$

$$L_6 = \{\sigma_1 w \sigma_2 \mid w \in \Sigma^*, \sigma_1, \sigma_2 \in \Sigma, \sigma_1 \neq \sigma_2\} \quad (6)$$

$$L_7 = \{w \mid |w| \bmod 3 = 1\} \quad (7)$$

$$L_8 = \left\{ w \mid \underbrace{\#a(w)}_{\text{number of appearances}} \bmod 2 = 1 \wedge \#b(w) \bmod 2 = 0 \right\} \quad (8)$$

$$\Sigma = \{a, b, c\} \quad , \quad \begin{aligned} L_9 &= \{a^i b^j c^k \mid i, j, k \geq 0\} \\ L_{10} &= \{a^i b^j c^k \mid i, j, k \geq 1\} \end{aligned} \quad (9)$$

$$F_A = \emptyset \text{ כל אס"ד } A \text{ כך ש-} F_A = \emptyset \quad (10)$$

$$. L = \{\varepsilon\} \text{ מקבל, וכל קלט הולך ל-"} \text{מלכודת"}. \quad (11)$$

$$. L = \Sigma^* \text{ מקבל, ומחזיר לעצמו.} \quad (12)$$

שפה רגולרית: כך שפה L אשר קיים אס"ד A כך ש- $L(A) = L$ נקראת שפה רגולרית. ניתן לומר כי L רגולרית אם קיים אס"ד A כך ש- $L(A) = L$.
לכן, כל השפות שראינו בתרגול על אס"ד הינן שפות רגולריות.

שיעור 3

אוטומט מכפלה

הרעיון הוא שאם L_1, L_2 שפות רגולריות מעל Σ ו- A_1, A_2 הם אס"ד המקבלים אותן בהתאמה, אזי ניתן לבנות אס"ד A כך שיתקיים $L(A) = L_1 \cap L_2$

נראה כיצד לבנות אוטומט A כזה

נרצה לבנות אס"ד A המחקה את פעולת A_1 ו- A_2 בו זמנית על כל מילת קלט. מצבי A יהיו הזוגות (q_{A_1}, q_{A_2}) כאשר $q_{A_1} \in Q_{A_1}$ ו- $q_{A_2} \in Q_{A_2}$. ממצב (q_{A_1}, q_{A_2}) ב- A נעבור למצב (p_{A_1}, p_{A_2}) ב- A עם קריאת הקלט $\sigma \in \Sigma$ אם מתקיימים התנאים הבאים:

$$\delta_{A_1}(q_{A_1}, \sigma) = p_{A_1} \text{ ו- } \delta_{A_2}(q_{A_2}, \sigma) = p_{A_2}.$$

המצבים המקבלים באוטומט יהיו כל הזוגות (q_{A_1}, q_{A_2}) המקיימים $q_{A_1} \in F_{A_1}$ וגם $q_{A_2} \in F_{A_2}$ (וגם: עבור חיתוך)

באופן פורמלי:

$$A_1 = \langle \sum, Q_{A_1}, q_{0_{A_1}}, F_{A_1}, \delta_{A_1} \rangle$$

$$A_2 = \langle \sum, Q_{A_2}, q_{0_{A_2}}, F_{A_2}, \delta_{A_2} \rangle$$

$$A = \langle \sum, Q_A, q_{0_A}, F_A, \delta_A \rangle$$

כאשר $Q_A = Q_{A_1} \times Q_{A_2}$, $q_{0_A} = (q_{0_{A_1}}, q_{0_{A_2}})$, $F_A = F_{A_1} \times F_{A_2}$ ולכל $\sigma \in \sum$ ולכל $q_{A_1} \in Q_{A_1}$, $q_{A_2} \in Q_{A_2}$ מתקיים:

$$\delta_A((q_{A_1}, q_{A_2}), \sigma) = (\delta_{A_1}(q_{A_1}, \sigma), \delta_{A_2}(q_{A_2}, \sigma))$$

דוגמה: $\sum = \{0, 1\}$, $L_1 = \{w \mid \#1(w) \bmod 2 = 1\}$, $L_2 = \{w \mid w \text{ contains substring } 000\}$

א) בנה אס"דים המקבלים את L_1, L_2

ב) בנה אס"ד מכפלה המקבל את $L_1 \cap L_2$

ג) בנה אס"ד מכפלה המקבל את $L_1 \cup L_2$

פתרון: במחברת

הערה: אוטומט מכפלה לא מבטיח מס' מינימלי של מצבים.

תרגיל: הוכח כי $L_1 = \{a^i b^i \mid i \geq 1\}$ לא רגולרית.

הוכחה: ראינו כי השפה $L = \{a^i b^i \mid i \geq 0\}$ לא רגולרית (בהרצאה). נניח בשלילה ש- L_1 רגולרית אזי מסגירות תחת איחוד: $L_1 \cup \{\varepsilon\}$ רגולרית, כי $\{\varepsilon\}$ רגולרית (ראינו כבר). אבל $L_1 \cup \{\varepsilon\} = \{a^i b^i \mid i \geq 0\} = L$ וזו סתירה לכך ש- L לא רגולרית.

טענה: השפות הרגולריות לא סגורות תחת איחוד אינסופי.

הוכחה: נניח בשלילה שכן. יהי $n \in \mathbb{N}$. נגדיר את השפה $L_n = \{a^n b^n\}$, מתקיים ש- L_n רגולרית. תהי $L = \bigcup_{n=0}^{\infty} L_n = \{a^n b^n \mid n \geq 0\}$ לפי ההנחה, L רגולרית, סתירה לכך ש- $\{a^n b^n \mid n \geq 0\}$ לא רגולרית.

טענה: השפות הרגולריות לא סגורות תחת חיתוך אינסופי.

הוכחה: נניח בשלילה שכן. יהי $n \in \mathbb{N}$. נתבונן ב- $L_n = \{a^n b^n\}$, מתקיים ש- L_n רגולרית. לפי סגירות תחת השלמה, גם השפה: $\overline{L_n} = \sum^* - \{a^n b^n\}$ רגולרית.

כעת נגדיר: $L' = \bigcap_{n=0}^{\infty} \overline{L_n} = \sum^* - \{\varepsilon\} - \{a^1 b^1\} - \{a^2 b^2\} - \dots = \sum^* - \{a^n b^n \mid n \geq 0\} = \overline{\{a^n b^n \mid n \geq 0\}}$ נגדיר: אזי לפי ההנחה שלנו, L' רגולרית, ולכן (שוב מסגירות תחת השלמה) $\overline{L'} = \{a^n b^n \mid n \geq 0\}$ רגולרית, וזו סתירה.

אוטומט סופי לא דטרמיניסטי (אסל"ד)

אסל"ד מוגדר ע"י חמישייה $\{\sum, Q, q_0, F, \delta\}$ כאשר: $\sum, Q, q_0, F, \delta$ כמו באסל"ד ופונקציית המעברים δ מוגדרת: $\delta: Q \times \sum \rightarrow 2^Q$ כלומר $\delta(q_1, \sigma) = \{q_1, \dots, q_r\}$ מציינן שאם נמצאים במצב q וקוראים את האות σ אז ניתן לעבור לכל אחד מהמצבים $\{q_1, \dots, q_r\}$.
ההרחבה של δ למילים מוגדרת כך: $\delta^*: Q \times \sum^* \rightarrow 2^Q$ ולכל $q \in Q$ ולכל $(\delta^*(q, \varepsilon) = \{q\})$ ולכל $q \in Q$,
$$\delta^*(q, w\sigma) = \bigcup_{q' \in \delta^*(q, w)} \delta^*(q', \sigma) \quad \sigma \in \sum, w \in \sum^*$$
 זו ההגדרה של אסל"ד ללא מעברי ε .

מעברי ε :

מעבר בו ניתן לעבור ממצב למצב ללא קריאת אות. כלומר, $\delta(q, \varepsilon) = p$ (כאשר $p \neq q$) הינו מעבר ε .
כל אסל"ד שבו מותרים מעברי ε ניתן להמיר לאסל"ד שקול ללא מעברי ε , לכן אנו נשתמש באסל"ד עם מעברי ε . כלומר, $\delta: Q \times (\sum \cup \{\varepsilon\}) \rightarrow 2^Q$ ובהתאם δ^* .
לכן אצלנו $\delta^*(q, w)$ יהיה אוסף כל המצבים שניתן להגיע אליהם מ- q לאחר קריאת המילה w תוך התחשבות בכל מעבר ε אפשרי.

שפת האוטומט

A אסל"ד. $L(A) = \{w \in \sum^* \mid \delta^*(q_0, w) \cap F \neq \emptyset\}$

כלומר: אוסף המילים שעבורן קיים חישוב (ריצה על האוטומט) המסתיים במצב מקבל.

מספר הבדלים בין אס"ד לאסל"ד

(1) באסל"ד לא חובה לתת פונקציה מעברים מלאה לכל מצב.

(2) באסל"ד ניתן להגדיר מעברי ε , בהם המעבר נעשה ללא קריאת תו.

(3) באסל"ד ניתן להגדיר עבור מצב ואות מסויימים מס' מעברים (כלומר מעברים למצבים שונים)

דוגמא:

$$L = \{w1 \mid w \in \Sigma^*\}$$

א) בנה אס"ד המקבל את L .

ב) בנה אסל"ד המקבל את L .

פתרון: במחברת.

שימו לב: לכל מילה בשפה צריך להיות חישוב שמסתיים במצב מקבל. ולכל מילה שלא בשפה לא קיים חישוב שמסתיים במצב מקבל.

דוגמאות במחברת

שיעור 4

משפט:

יהי A אסל"ד, אזי קיים A' אס"ד כך ש- $L(A') = L(A)^-$. ראינו בהרצאה כי ניתן לבנות A' כנ"ל שקבוצת מצביו היא קבוצת החזקה של מצבי A .

הערה: בנייה זו אינה מבטיחה מס' מינימלי של מצבים ויש באוטומט החזקה $2^{|Q_A|}$ מצבים שזה חסם עליון למס' המצבים.

דוגמה

נתון האסל"ד הבא: ציור במחברת

צייר אס"ד שקול לו.

$$Q_A = \{q_0, q_1, q_2\}$$

באסל"ד שציירנו יש 6 מצבים. שאר מצבי האוטומט החזקה: $\{q_1, q_2\}, \{q_0, q_1\}$ לא היו נחוצים משום שלא ניתן היה להגיע אליהם מהמצב ההתחלתי.

ביטויים רגולריים

הגדרה: אחת השיטות לייצוג של שפות, היא ע"י ביטוי רגולרי. לכל ביטוי רגולרי מתאימה שפה בהינתן א"ב Σ , ביטוי רגולרי הוא מילה מעל הא"ב $\Sigma \cup \{\emptyset, \cup, (,), \cdot, *\}$ השייך לשפה המוגדרת כך:

(1) $\emptyset$ וכל איבר ב- Σ הוא ביטוי רגולרי

(2) אם α ביטוי רגולרי, אז גם α^* הוא ביטוי רגולרי

(3) אם α, β הם ביטויים רגולריים אז גם $\alpha \cup \beta$ ביטוי רגולרי

(4) אם α, β הם ביטויים רגולריים אז גם $\alpha \cdot \beta$ ביטוי רגולרי

סדר פעולות

*- קליני קודמת לשרשור, שרשור קודם לאיחוד. כך נוכל לצמצם את מופעי הסוגריים בביטוי. דוגמה:
 $((a(b^*)) \cup (b(a^*))) \equiv (ab^* \cup ba^*)$
 שרשור נרשום בפשטות כרצף של תווים. דוגמה: $r = (0 \cdot 1 \cdot 0) \equiv 010$

סמנטיקה של ביטויים רגולריים:

(1) $L(r) = \emptyset$ משמעו $r = \emptyset$

(2) $L(r) = \{\sigma\}$ משמעו $r = \sigma$

(3) $L(r) = L(r_1) \cup L(r_2)$ משמעו $r = (r_1 \cup r_2)$

(4) $L(r) = L(r_1) \cdot L(r_2)$ משמעו $r = (r_1 \cdot r_2)$

(5) $L(r) = (L(r_1))^*$ משמעו $r = r_1^*$

דוגמה

(1) $\Sigma = \{0, 1\}$, וביטוי רגולרי $r = (0 \cup 10) \cdot (01 \cup 11)$ זהו השפה $L(r) = \{001, 011, 1001, 1011\}$

(2) $L = \{a^i b^j c^k \mid i, j, k \geq 0\}$, ביטוי רגולרי עבור L : $(a^* b^* c^*)$

משפט

L שפה רגולרית אם ורק אם קיים ביטוי רגולרי r כך שמתקיים $L(r) = L$

הוכחה:

$$L(A) = L \text{ ש-} A \text{ (} \implies \text{) קיים אס"ד } A$$

$$L(A') = L \text{ ש-} A' \text{ (} \Longleftarrow \text{) קיים אסל"ד } A'$$

תרגיל

ההארה שגם השפה L^R רגולרית. נגדיר $L^R = \{w^R \mid w \in L\}$ w^R מציין $(reverse(w))$ ההארה שגם השפה L^R רגולרית.

הוכחה: יהי r ביטוי רגולרי כך ש- $L(r) = L$ (קיים כזה, כי L רגולרית). נראה כי לשפה L^R יש ביטוי רגולרי r' כך ש- $L(r') = L^R$ ונסיק מכך שגם L^R רגולרית. הוכיח זאת באינדוקציה שלמה על האורך של r .

בסיס $|r| = 1$, מכאן ש- $r = \emptyset$ או $r = \sigma$. בשני המקרים $L^R = L$ ומכאן $r' = r$ הוא ביטוי רגולרי המתאר את L^R .

הנחת האינדוקציה נניח שלכל $1 \leq k < n$, אם L שפה עבודה קיים ביטוי רגולרי r באורך k , כך ש- $L(r) = L$, אזי קיים ב"ר r' כך ש- $L(r') = L^R$.

הוכחת צעד האינדוקציה יהי r ביטוי רגולרי $L = L(r)$ ו- $|r| = n$. מכיוון ש- $|r| = n > 1$ נפריד לשלושה מקרים. (1) מהצורה: $r = (r_1 \cup r_2)$ (2) מהצורה: $r = (r_1 \cdot r_2)$ (3) מהצורה: $r = (r_1^*)$

מקרה 1: נבחר r_1, r_2 מרכיבים את r ושניהם באורך 1 לפחות ואילו r מכיל תווים נוספים. מכאן, $|r_1|, |r_2| < n$. לפי הנחה קיימים ביטויים רגולריים r'_1, r'_2 עבורם מתקיים $L(r'_1) = L(r_1)^R, L(r'_2) = L(r_2)^R$. נסמן $r' = (r'_1 \cup r'_2)$.

$$L(r') = L^R \text{ טענה:}$$

הוכחה: כדי להראות $L(r') = L^R$, נראה כי $w \in L \iff w^R \in L(r')$.

$$w \in L \iff w \in L(r) = L(r_1 \cup r_2) \underbrace{=}_{\text{semantics}} L(r_1) \cup L(r_2) \iff w \in L(r_1) \vee w \in L(r_2) \underbrace{\iff}_{\text{assumption}}$$

$$w^R \in L(r'_1) \vee w^R \in L(r'_2) \iff w^R \in L(r'_1) \cup L(r'_2) = L(r'_1 \cup r'_2) = L(r')$$

מקרה 2: כמו במקרה הקודם, לפי הנחה קיימים r'_1, r'_2 עבורם מתקיים $L(r'_1) = L(r_1)^R, L(r'_2) = L(r_2)^R$. נסמן $r' = (r'_2 \cdot r'_1)$ (נשים לב כי הפכנו את הסדר!).

טענה: $L(r') = L^R$

הוכחה: כדי להראות $L(r') = L^R$, נראה כי $w \in L \iff w^R \in L(r')$.

$$w \in L \iff w \in L(r) = L(r_1 \cdot r_2) \underset{\text{semantics}}{=} L(r_1) \cdot L(r_2) \iff$$

$$w = w_1 \cdot w_2 : w_1 \in L(r_1) \wedge w_2 \in L(r_2) \underset{\text{assumption+reverse}}{\iff}$$

$$w^R = w_2^R \cdot w_1^R : w_1 \in L(r'_1) \wedge w_2^R \in L(r'_2) \iff w^R \in L(r'_2) \cdot L(r'_1) = L(r'_2 r'_1) = L(r')$$

מקרה 3: כמו במקרה הקודם, לפי הנחה קיים ביטוי רגולרי r'_1 עבורו מתקיים $L(r'_1) = L(r_1)^R$. נסמן $r' = (r'_1)^*$.

טענה: $L(r') = L^R$

הוכחה: כדי להראות $L(r') = L^R$, נראה כי $w \in L \iff w^R \in L(r')$.

$$w \in L \iff w \in L(r) = L(r_1 \cdot r_2) \underset{\text{semantics}}{=} (L(r_1))^* \iff$$

$$w = w_1 \cdot w_2 \cdot \dots \cdot w_k : \forall i (1 \leq i \leq k) w_i \in L(r_1) \underset{\text{assumption+reverse}}{\iff}$$

$$w^R = w_k^R \dots w_1^R, \forall i (1 \leq i \leq k) w_i^R \in L(r'_1) \iff w^R \in (L(r'_1))^* = L(r'_1)^* \underset{\text{semantics}}{=} L(r')$$

תרגיל

(א) מצא ב"ר המגדיר את שפת האסל"ד: (במחברת)

(ב) צייר אסל"ד המקבל את השפה $L((101 \cup 010)^*)$

תרגילים

כתבו ביטוי רגולרי המציין את השפות הבאות מעל $\Sigma = \{0, 1\}$

(א) שפת המילים המכילות 10 כתת-מילה.

פתרון: $(1 \cup 0)^* 10 (1 \cup 0)^*$

(ב) שפת המילים המסתיימות ב-11.

פתרון: $(1 \cup 0)^* 11$

(ג) שפת המילים שאורכן זוגי.

פתרון: $((1 \cup 0) \cup (1 \cup 0))^*$

(ד) שפת המילים בהן יש בדיוק 1 פעם אחת.

פתרון: $0^* 10^*$

(ה) השפה $L = \{w \mid \#a(w) = \#b(w) \wedge \text{for all } x = \text{prefix of } w: |\#a(x) - \#b(x)| \leq 1\}$ (מעל $\sum$) $\{a, b\}$

פתרון: $(ab \cup ba)^*$

שיעור 5

לימת הניפוח

אם L רגולרית, אזי קיים קבוע N כך שלכל $w \in L$ שמקיימת $|w| \geq N$ קיימת חלוקה למילים $x, y, z \in \Sigma^*$ כך ש: $w = xyz$ וזומתקיימים:

$$1. |y| \geq 1 \text{ ו- } y \neq \varepsilon$$

$$2. |xy| \leq N$$

$$3. xy^i z \in L \text{ לכל } i \geq 0$$

משפט: L רגולרית $\iff L$ מקיימת את תנאי למת הניפוח

נשים לב כי הכיוון השני לאו דווקא נכון.

הערה: לימת הניפוח תופיע במבחן.

רעיון הוכחת אי רגולריות של שפה L ע"י למת הניפוח

נניח בשלילה ש- L רגולרית ונרצה להראות סתירה ללמת הניפוח. כדי להראות סתירה, יש להראות כי: לכל N טבעי, קיימת מילה $w \in L$ כך ש- $|w| \geq N$ שלכל חלוקה אפשרית $w = xyz$ קיים $i \geq 0$ שעבורו $xy^iz \notin L$.

תרגיל: הוכח שהשפה $L = \{w \mid \#a(w) = \#b(w)\}$ איננה רגולרית (בעזרת למת הניפוח)

פתרון: נניח בשלילה כי L רגולרית. יהי n הקבוע המובטח מלמת הניפוח. נבחר את המילה $w = a^n b^n$. מתקיים $w \in L$ ולכן לפי למת הניפוח קיים, פירוק $w = xyz$ כך ש:

$$1. y \neq \varepsilon$$

$$2. |xy| \leq n$$

$$3. xy^iz \in L \text{ לכל } i \geq 0$$

מתנאי 2 עולה כי xy כלול ברישא a^n . (הסבר: $|a^n| = |b^n| = n$ ולכן $|a^n b^n| = 2n$) ולפי תנאי 1, $y \neq \varepsilon$ כלומר קיימים s, t כך ש- $0 \leq s, 1 \leq t \leq n$ שעבורם מתקיים $s + t \leq n$. עבור אותם s, t יתקיים

$$x = a^s \quad y = a^t \quad z = a^{n-s-t} b^n$$

תנאי 3 קובע שלכל $i \geq 0$ מתקיים $xy^iz \in L$. נבחר $i = 0$ ונקבל $xy^0z = xz = a^s a^{n-s-t} b^n = a^{n-t} b^n$. ההנחה $a^{n-t} b^n \in L$ אך $n - t \neq n$ כי $t \geq 1$ ולכן $a^{n-t} b^n \notin L$, סתירה. לכן לא מקיימת את למת הניפוח ולכן L לא רגולרית.

תרגיל: הוכח כי $L = \{w \mid w = w^R\}$ לא רגולרית (נזכיר $w^R = \text{reverse}(w)$)

כיצד לפתור? נניח בשלילה ש- L רגולרית. יהי n הקבוע שקיומו מובטח מלמת הניפוח. נבחר $w = \underbrace{a^n}_{\text{ז"א}}, z = \underbrace{a^n}_{\text{ז"א}}$ כך ש- $1 \leq |y|, |xy| \leq n$. נבחר $i = ?$ וע"פ למת הניפוח $xy^iz \in L$ ונרצה להראות כי $xy^iz \notin L$ ובכך להגיע לסתירה.

ראשית נדגים בחירה לא טובה של w :

אם היינו בוחרים $w = a^n a^n$ אז כל פירוק יהיה מהצורה $x = a^s, y = a^t, z = a^{n-s-t} a^n$ עבור $t \geq 1, s \geq 0$. היינו בוחרים למשל $i = 0$: $xy^0z = a^s a^{n-s-t} a^n = a^{n-t} a^n \in L$.

לא טוב! כל בחירה של i לא הייתה עוזרת כאן.

נעת נדגים בחירה טובה של w :

נבחר $w = a^n b a^n$. ואז כל פירוק יהיה מהצורה $x = a^s, y = a^t, z = a^{n-s-t} b a^n$ עבור $t \geq 1, s \geq 0$. נבחר $i = 2$ ואז מתקיים: $xy^2z = a^s a^{2t} a^{n-s-t} b a^n = a^{n+t} b a^n$. אבל $w' \in L$ מההנחה אבל $w' \neq w^R$ (כי $n + t \neq n$) ולכן $w' \notin L$, והגענו לסתירה.

תרגיל:

הוכח כי $L = \{a^{2^j} \mid j \geq 0\}$ (כאשר: $\{a\}$)

פתרון: L היא שפת המילים מעל $\{a\}$ שאורכן הוא חזקה של 2. למשל a^8, a^{32} וכו'. נניח בשלילה כי L רגולרית ויהי n הקבוע שקיומו מובטח מלמת הניפוח. נבחר $w = a^{2^n}$. מתקיים כי $w \in L$, כמו כן $|w| > n$ (גדול ממש מכיוון ש- $2^n > n$ לכל $n \in \mathbb{N}$). לפי למת הניפוח קיים פירוק $w = xyz$ כך ש: $|xy| \leq n, |y| \geq 1, xy^iz \in L$ לכל $i \geq 0$.

עבור $i = 2$ נקבל: $|xy^2z| = 2^n + |y|$ (כי $|w| = |xyz| = 2^n$) והוספנו עוד $|y|$ נשים לב כי: $1 \leq |y| \leq n$ (כי $|xy| \leq n$ וגם $|y| \geq 1$) מכאן מקבלים:
 $2^n < |xy^2z| < 2^{n+1}$ כלומר $2^n < 2^n + 1 \leq |xy^2z| \leq 2^n + n < 2^n + 2^n = 2^{n+1}$
לכן $xy^2z \notin L$ איננו מהצורה a^{2^j} עבור $j \in \mathbb{N}$ כלומר $xy^2z \notin L$ והגענו לסתירה.

תרגיל: הוכח כי $L = \{a^p \mid p \text{ is prime}\}$ לא רגולרית.

הוכחה: נניח בשלילה ש- L רגולרית. יהי n הקבוע שקיומו מובטח מלימת הניפוח. נבחר $w = a^p$ כאשר $n < p$ (יש p כזה, כי יש אינסוף מס' ראשוניים). מתקיים $w \in L, |w| \geq n$ ועבור פירוק $w = xyz$ הוא יהיה מהצורה: $a^p = \underbrace{a^s}_x \underbrace{a^t}_y \underbrace{a^{p-s-t}}_z$ כאשר $0 \leq s$, $1 \leq t$ ו- $s+t \leq n$.

נבחר $i = p+1$: $a^s a^{(p+1)t} a^{p-s-t} = a^s a^{pt} a^t a^{p-s-t} = a^{pt} a^p = a^{(t+1)p}$ נשים לב ש- $p \geq 2$ מכיוון שהוא ראשוני, ו- $t+1 \geq 2$ ומכאן $(t+1)p$ אינו ראשוני (הוא מכפלה של 2 מספרים: p ו- $t+1$). לכן, $a^{(t+1)p} \notin L$, והגענו לסתירה.

שאלה: תהי L השפה הכוללת את המילים מעל אותיות הא"ב העברי כך שאף אות לא מוספיעה יותר מ-3 פעמים. קבע האם L מקיימת את תנאי למת הניפוח והוכח טענתך.

תשובה: אורך מילה ב- L חסום ע"י הקבוע $66 = 22 \cdot 3$. אם נבחר את הקבוע n מלמת הניפוח להיות 67, נוכל לראות שהשפה מקיימת את תנאי הלמה באופן ריק. למעשה, עבור כל שפה סופית נוכל לבחור את n להיות אורך המילה הארוכה ביותר בשפה + 1.

נראה דוגמה לשפה לא רגולרית המקיימת את למת הניפוח

$$L = \{b^j a^p \mid j \geq 0, p \text{ is prime}\} \cup \{a\}^*, \Sigma = \{a, b\}$$

נראה ש- L ניתנת לניפוח: ראשית נקבע מהו n : כל $n \geq 2$ יתאים. כל מילה $w \in L$ באורך n לפחות היא מאחת הצורות הבאות:

$$1. w = b^j a^p \text{ עבור } j \geq 1, p \text{ ראשוני, } j+p \geq n$$

$$2. w = a^k \text{ עבור } k \leq n \text{ (גם מילים מהצורה } b^0 a^p \text{ כלולות כאן)}$$

$$\text{במקרה 1 אפשר לפרק את } w \text{ להיות מהצורה } z = b^{j-1} a^p, y = b, x = \varepsilon,$$

$$\text{במקרה 2 אפשר לפרק את } w \text{ להיות מהצורה } z = a^{k-1}, y = a, x = \varepsilon,$$

קל לראות ששני הפירוקים מקיימים את תנאי הלמה. כמו כן ניתן להוכיח ש- L לא רגולרית.

שיעור 6

מחלקות שקילות ב- Σ^*

תהיינה שתי מחרוזות $x, y \in \Sigma^*$

נגדיר סיפא מפרידה: $z \in \Sigma^*$, תקרא סיפא מפרידה בין x ל- y ביחס לשפה L אם מתקיים:

$$(xz \in L \wedge yz \notin L) \vee (xz \notin L \wedge yz \in L)$$

נגדיר יחס $\equiv_L$: נאמר כי $x \equiv_L y$ אם מתקיים $xz \in L \iff yz \in L$ לכל $z \in \Sigma^*$ (לכל סיפא). למעשה, תנאי זה אומר כי אין סיפא מפרידה בין x ל- y .

קל לוודא שהיחס $\equiv_L$ הוא יחס שקילות: הוא רפלקסיבי, סימטרי וטרנזיטיבי.

תכונות היחס $\equiv_L$

- היחס $\equiv_L$ מגדיר חלוקה של כל Σ^* למחלקות שקילות כך שלכל זוג מילים מאותה מחלקת שקילות אין סיפא מפרידה ולכל שתי מילים ממחלקות שונות יש סיפא מפרידה.
- כל מילה מוכלת ב- L או ב- $\bar{L}$. כלומר, לא יתכן כי $w_1 \in L, w_2 \notin L$ ושהמילים w_1, w_2 מאותה מחלקת שקילות. (כאן $z = \varepsilon$ תהיה סיפא מפרידה)

הגדרה: $Rank(L)$ - מס' מחלקות השקילות ביחס $\equiv_L$

משפט (Myhill – Nerod):

1. L שפה רגולרית $\iff Rank(L)$ הוא מס' סופי.
2. תהי L שפה עבודה מתקיים $Rank(L) = n$ אז קיים אס"ד בעל n מצבים המקבל את L , ולא קיים אס"ד עם פחות מ- n מצבים המקבל את L .

תכונות היחס $\equiv_L$ (המשך):

- איחוד מחלקות השקילות שמגדיר $\equiv_L$ הוא Σ^*
- כל שתי מילים שנמצאות באותה מחלקת שקילות "יעברו יחד" לאותה מחלקת שקילות. כלומר: אם w_1, w_2 באותה מחלקת שקילות, אזי w_1z, w_2z גם באותה מחלקת שקילות, לכל סיפא $z \in \Sigma^*$
- מחלקת שקילות של איבר $x \in \Sigma^*$ נסמן ב- $[x]$. $[x] = \{y \mid x \equiv_L y\}$ ז"א: $[x]$ היא קבוצת כל האיברים y כך ש: $x \equiv_L y$

דוגמאות

נניח $\Sigma = \{a, b\}$

(1) תהי $L = \{a, aa, aaa\}$. מה יהיו מחלקות השקילות של $\equiv_L$?

תשובה: $S_1 = \{\varepsilon\}, S_2 = \{a\}, S_3 = \{aa\}, S_4 = \{aaa\}, S_5 = \Sigma^* - \{\varepsilon, a, aa, aaa\}$

קל להראות שהמחלקות זרות בזוגות ואיחודם הוא Σ^* . נראה שלכל שתי מילים מ- S_5 אין סיפא מפרידה: נשים לב שכל המילים שמכילות b -ים או יותר מ-3 a -ים נמצאות ב- S_5 , משום שלכל שתי מילים $w_1, w_2 \in S_5$ מתקיים שלכל $z \in \Sigma^*$: $w_1z, w_2z \notin L$. כלומר אין להן סיפא מפרידה ולכן הן באותה מחלקה. נראה שלמילים ממחלקות שונות יש סיפא מפרידה. נראה $a \not\equiv_L \varepsilon$: נבחר $z = aaa$ ונקבל: $yz = aaa \in L$ ואילו $az = aaaa \notin L$. כלומר z סיפא מפרידה ולכן $a \not\equiv_L \varepsilon$. באופן דומה קל להוכיח שכל אחת מהמילים: ε, a, aa, aaa אינה שקולה לרעותה. כמו כן אף אחת מהמילים הללו לא שקולות למילים מ- S_5 . ניתן להראות אם נבחר למשל $b \in S_5$ (מספיק לבחור נציג מ- S_5 אחרי שהראינו ש- S_5 היא מחלקת שקילות). עבור $z = \varepsilon$: $bz = b \notin L$ ואילו $az = a$, $aaaz = aaa$, $aaaz = aa$, $aaaz = a$ וכולן ב- L . כדי להראות ש- ε לא שקולה למילים מ- S_5 נבחר $z = a$.

כאשר מבקשים מאיתנו למצוא במפורש את מחלקות השקילות לפי היחס $\equiv_L$ יש להראות:

1. המחלקות זרות בזוגות (אין מילה ב- Σ^* ששייכת ליותר ממחלקה אחת)

2. איחוד המחלקות הוא Σ^*

3. לכל שתי מילים מאותה מחלקה אין סיפא מפרידה ביחס ל- L .

4. לכל שתי מילים ממחלקות שונות יש סיפא מפרידה ביחס ל- L .

(2)

1. מהן מחלקות השקילות של $\equiv_L$ עבור $L = \{w \mid |w| \bmod 3 = 0\}$

2. בנה אס"ד מינימלי ל- L .

$$S_1 = \{w \mid |w| \bmod 3 = 0\}$$

תשובה: $S_2 = \{w \mid |w| \bmod 3 = 1\}$ קל לראות (יש להסביר) כי לכל שתי מילים מאותה מחלקה אין סיפא

$$S_3 = \{w \mid |w| \bmod 3 = 2\}$$

מפרידה ולכל שתי מילים ממחלקות שונות יש סיפא כזו, כמו כן ברור כי $S_1 \cup S_2 \cup S_3 = \Sigma^*$ (שהן זרות בזוגות).

על מנת לבנות אס"ד ממחלקות השקילות: נבנה מצב לכל מחלקת שקילות. המצב ההתחלתי יהיה $[\varepsilon]$ (המחלקה ש- ε בה) ופונקציית המעברים תהיה $\delta([x], a) = [xa]$. נשים לב כי $[xa]$ זו מחלקת שקילות ולא משנה איזה נציג נבחר ממנה.

שרטוט במחברת

אצלנו $\underbrace{[aa] = [ab]}_{\text{same department}}$. שמות המצבים מתייחסים למחלקה כולה, למשל אצלנו $[ab]$ ו- $[aa]$ מתייחס לאותו מצב.

(3) מהן מחלקות השקילות של $\equiv_L$ עבור $L = \{w \mid \#a(w) = \#b(w)\}$:

תשובה: נשים לב שכל מילה ב- $\sum^*$ נמצאת בדיוק באחת משתי המחלקות. כלומר, הן זרות בזוגות ואיחודן הוא $\sum^*$.

$$S_k = \{w \mid \#a(w) - \#b(w) = k\} \quad k \geq 0$$

$$T_k = \{w \mid \#b(w) - \#a(w) = k\} \quad k > 0$$

נראה שלכל 2 מילים מאותה מחלקה אין סיפא מפרידה: לכל $0 < k$ אם $x, y \in S_k$ אזי עבור $z \in T_k$ נקבל: $yz \in L \wedge xz \in L$. אחרת, נקבל $xz \notin L \wedge yz \notin L$. לכן, כל המילים ב- S_k עומדות ביחס $\equiv_L$ זו לזו. הטענה נכונה גם עבור S_0 וגם על T_K , $k > 0$.

נראה שלכל 2 מילים ממחלקות שונות יש סיפא מפרידה: נבצע חלוקה למקרים:

$$1. \quad i \neq j \quad y \in S_j, x \in S_i$$

$$2. \quad i \neq j \quad y \in T_j, x \in T_i$$

$$3. \quad y \in T_j, x \in S_i \quad \text{עבור } j > 0, i \geq 0$$

מקרה 1: יהיו $x \in S_i, y \in S_j, i \neq j$ נראה כי $x \not\equiv_L y$: נניח בה"כ $j < i$ וניקח $z = b^j$. אזי $yz \in S_0 = L$ אבל $xz \notin L$ כי $xz \in S_{i-j}$ ו- $i-j \neq 0$ ($i > j$)

מקרה 2: סימטרי למקרה 1.

מקרה 3: יהיו $x \in S_i, y \in T_j, i \geq 0, j > 0$. אזי עבור $z = b^j$ נקבל $xz \in L \wedge yz \notin L$ ולכן z סיפא מפרידה, כלומר $x \not\equiv_L y$.

תרגיל (חשוב): תהי $L = \{w_1cw_2 \mid w_1, w_2 \in \{a, b\}^* \wedge \#a(w_1) = \#a(w_2)\}$ מעל $\Sigma = \{a, b, c\}$. הוכח באמצעות משפט מייהל נרוד כי L לא רגולרית.

פתרון: נרצה להראות שיש מס' אינסופי של מחלקות שקילות, ז"א $Rank(L)$ אינסופי. קיימות אינסוף מילים ב- $\sum^*$ הניתנות להפרדה (ביחס ל- L) בזוגות.

נתבונן במילים מהצורה $c, ac, aac, aaac, \dots$ כלומר מילים מהצורה a^nc עבור $n \geq 0$. נתבונן בזוג מילים שונות a^ic ו- a^jc , כאשר $(i \neq j)$

a^i למשל היא סיפא מפרידה עבורן, שכן $a^ic \in L$ ו- $a^jca^i \notin L$ אם כן, כל מילה מהצורה a^n : $n \geq 0$ היא ממחלקת שקילות שונה. כלומר, יש אינסוף מילים כאלה ולכן יש אינסוף מחלקות שקילות ולכן L איננה רגולרית.

אבחנה: בהינתן שפה L , אם נוכל למצוא אינסוף מילים מ- $\sum^*$ כך שכל זוג מילים ניתן להפרדה לפי היחס $\equiv_L$, אזי L אינה רגולרית שכן במקרה זה יש אינסוף מחלקות שקילות.

שיעור 7

דקדוקים

דקדוק פורמלי ע"י רביעייה (Σ, V, S, Π) כאשר: Σ הינו א"ב, נקרא לאיברי Σ טרמינלים (אותיות סופיות). V - קבוצה סופית ולא ריקה של משתנים דקדוקיים (אותיות לא סופיות), נסמן אותן באותיות גדולות $(X, V_1, \dots)$. $S \in V$ משתנה התחלתי. Π - קבוצה סופית של כללי גזירה (הנקראים גם "הפקות") מהצורה $\{g_i \rightarrow h_i\}_{i=1}^l$ כאשר $g_i, h_i \in (V \cup \Sigma)^*$

הגדרה:

1. עבור $\varphi_1, \varphi_2 \in (V \cup \Sigma)^*$ נאמר כי φ_2 נגזר ישירות מ- φ_1 ונסמן $\varphi_1 \xRightarrow{\Pi} \varphi_2$ אם קיימת הפקה

כאשר $\varphi_1 = r \cdot g_i \cdot s$ ו- $\varphi_2 = r \cdot h_i \cdot s$ כך ש: $g_i \rightarrow h_i$, כלומר ניתן לעבור מ- φ_1 ל- φ_2 ע"י הפעלה אחת של כלל מ- Π

2. נסמן $\varphi_1 \xRightarrow{*} \varphi_2$ אם ניתן להגיע מ- φ_1 ל- φ_2 ע"י הפעלת מס' סופי של כללים מ- Π בזה אחר זה.
 $\varphi_1 \xRightarrow{\Pi} U_1 \xRightarrow{\Pi} \dots \xRightarrow{\Pi} \varphi_2 \in (\Sigma \cup V)^*$

שפת הדקדוק: יהי הדקדוק $G = (\Sigma, V, S, \Pi)$, השפה המתקבלת מ- G היא $L(G) = \left\{ w \in \Sigma^* \mid S \xRightarrow{*} w \right\}$

דקדוק רגולרי: הוא דקדוק שכללי הגזירה ב- Π הם מהצורה הבאה בלבד:

$$\begin{cases} U \rightarrow aW & U, W \in V, a \in \Sigma \\ U \rightarrow a & U \in V, a \in \Sigma \\ U \rightarrow \varepsilon \end{cases}$$

(1) אות סופית ואחריה אות לא סופית
(2) אות סופית
(3) ε - האות הריקה

כלומר: הצד הימני מוגבל לאופציות:

דוגמה לדקדוק רגולרי

$$L(G_1) = \{ab\}^* a \quad \text{כאן: } G_1 : \begin{matrix} S \rightarrow aB|a \\ B \rightarrow bA \end{matrix}$$

2. דקדוק שקול ל- G_1 בעל שני משתנים בלבד: $G_2 : \begin{matrix} S \rightarrow aB|a \\ B \rightarrow bS \end{matrix}$, מתקיים $L(G_1) = L(G_2)$

נדגים כיצד המילה $ababa$ נגזרת מ- G_1, G_2 :

$$S \Rightarrow aB \Rightarrow abA \Rightarrow abaB \Rightarrow ababA \Rightarrow ababa \quad (\alpha)$$

$$S \Rightarrow aB \Rightarrow abS \Rightarrow abaB \Rightarrow ababS \Rightarrow ababa \quad (\text{ב})$$

דוגמה לדקדוק לא רגולרי

$$S \rightarrow aSa|bSb|SS|a|b|\varepsilon|ab|aS$$

משפט: L שפה רגולרית $\iff$ קיים דקדוק רגולרי G כך ש: $L(G) = L$

הערה: קיימים דקדוקים לא רגולריים שהשפות שהם יוצרים הן רגולריות (יש גם דקדוק רגולרי שיוצר אותן)

למשל: $S \rightarrow aSa|\varepsilon$ זה דקדוק לא רגולרי שיוצא את $\{aa\}^*$ רגולרית $(aa)^*$ ביטוי רגולרי עבורה.

שפות חסרות הקשר

דקדוק חסר הקשר: זהו דקדוק בו כל כללי הגזירה של Π הם מהצורה $X \rightarrow \beta$ כאשר $X \in V$ ו- $\beta \in (\sum \cup V)^*$ (הערה: הכלל $X \rightarrow \varepsilon$ מותר)

הגדרה: שפה L היא חסרת הקשר $\iff$ קיים דקדוק חסר הקשר G כך ש: $L(G) = L$

נשים לב שכל דקדוק רגולרי הוא גם דקדוק חסר הקשר ומכאן שכל שפה רגולרית היא גם שפה חסרת הקשר (אבל לא להיפך!)

דוגמאות לדקדוק חסר הקשר

$G: S \rightarrow aSa|bSb|\varepsilon$ איזו שפה זאת? $L(G) = \{ww^R | w \in \{a,b\}^*\}$. ניתן להראות ש- $L(G)$ לא רגולרית ע"י למת הניפוח. נראה גזירה עבור $w = abba$: $S \Rightarrow aSa \Rightarrow abSba \Rightarrow abba$

הגדרה: עץ גזירה עבור דקדוק ח"ה $G = (\sum, V, S, \Pi)$ הינו עץ סדור המקיים:

1. כל צומת מסומן באות מתוך $V \cup \sum \cup \{\varepsilon\}$
2. השורש מסומן ב- S (המשתנה ההתחלתי)
3. הסימון של כל צומת פנימי (איננו עלה) הוא מתוך V
4. צומת המסומן ב- ε הוא בן יחיד
5. אם צומת פנימית מסומן ב- $A \in V$ ואם לבניו יש סימונים (לפי הסדר משמאל לימין) $X_1 X_2 \dots X_n$ אזי $A \rightarrow X_1 \dots X_n$ הינו כלל של Π

תרגיל

1. מהי שפת הדקדוק $G: S \rightarrow aSb|ab$
2. צייר עץ גזירה עבור $w = a^3b^3$

תשובה

$$1. L(G) = \{a^n b^n \mid n \geq 1\}$$

2. במחברת

תרגילים: $\Sigma = \{a, b\}$ תן דקדוק חסר הקשר עבור השפות הבאות מעל

1. שפת המילים המתחילות ברצף ab

2. שפת המילים באורך זוגי

3. שפת המילים באורך אי-זוגי

$$4. L = \{w \mid \#b(w) \bmod 2 = 0\}$$

תשובות:

$$1. \begin{aligned} S &\rightarrow abX \\ X &\rightarrow aX \mid bX \mid \varepsilon \end{aligned}$$

$$2. \begin{aligned} S &\rightarrow AAS \mid \varepsilon \text{ או } S \rightarrow abS \mid baS \mid aaS \mid bbS \mid \varepsilon \\ A &\rightarrow a \mid b \end{aligned}$$

3. עבור אי זוגי נחליף את הכלל $S \rightarrow \varepsilon$ בתשובות ל-2 ב- $a \mid b$

$$4. \begin{aligned} S &\rightarrow AbAbS \mid A \\ A &\rightarrow aA \mid \varepsilon \end{aligned}$$

נשים לב: כל השפות הנ"ל הן רגולריות והראינו עבורן דקדוקים לא רגולריים (חסרי הקשר) אבל לכל שפה מהשפות הללו קיים גם דקדוק רגולרי שיוצר אותן.

$$\begin{aligned} S &\rightarrow aB \mid cB \mid cA \mid a \mid b \mid \varepsilon \\ A &\rightarrow aA \mid \varepsilon \mid b \\ B &\rightarrow bB \mid aB \mid cB \mid \varepsilon \end{aligned}$$

השפה הנוצרת ע"י הדקדוק היא רגולרית, משום שהדקדוק הוא רגולרי.

תרגיל: תן דקדוק ח"ה לשפות הבאות מעל $\Sigma = \{a, b\}$:

$$1. \{a^n b^m \mid n, m \geq 0, n \neq m\}$$

$$2. \{a^n b^m \mid n, m \geq 0, n \leq 3m\}$$

תשובות:

$$\begin{aligned}
 & S \rightarrow aSb \rightarrow aaSbb \rightarrow aaaAbb \rightarrow :w = aaabb \text{ נראה לדוגמה גזירה עבור } \\
 & \begin{array}{l} S \rightarrow aSb|aA|bB \\ A \rightarrow aA|\varepsilon \\ B \rightarrow bB|\varepsilon \\ \quad \quad \quad aaabb \end{array} \quad 1. \\
 & S \rightarrow aaaSb|aaSb|aSb|Sb|\varepsilon \quad 2.
 \end{aligned}$$

שיעור 8

הצורה הנורמלית של חומסקי

כל שפה חסרת הקשר, שאינה מכילה את ε אפשר ליצור אותה על ידי דקדוק שכל כלליו מהצורה:

$$\begin{array}{l} V_0 \rightarrow V_1 V_2, \quad V_0, V_1, V_2 \in V \\ V_0 \rightarrow a, \quad a \in \Sigma \end{array}$$

דקדוק כזה נקרא דקדוק בצורה הנורמלית של חומסקי.

תזכורת: אם G' דקדוק חסר הקשר אזי קיים דקדוק חסר הקשר G'' חיובי ומסתעף.

$$L(G) = L(G') \cup \{\varepsilon\} \quad \text{כך ש:} \quad \begin{array}{l} \text{(חיובי)} \quad X \rightarrow \varepsilon, X \in V \\ \text{(מסתעף)} \quad X \rightarrow Y, X, Y \in V \end{array}$$

כעת נראה אלגוריתם מעבר מדקדוק חסר הקשר G כך ש: $\varepsilon \notin L(G')$ לדקדוק חסר הקשה בצורה הנורמלית של חומסקי:

תחילה נוודא ש- G' הינו דקדוק ח"ה חיובי ומסתעף. אם לא, נהפוך אותו לכזה (ראינו בהרצאה כיצד מעבר כזה הוא אפשרי) כעת נבנה דקדוק G'' באופן הבא:

אם $A \rightarrow a, a \in \Sigma$ כלל ב- G' אז הוא גם כלל ב- G'' . (כלל = הפקה). נותרו כללים מהצורה $(*) := A \rightarrow X_1 X_2 \dots X_t$ כאשר $2 \leq t$ ו- $X_i \in \Sigma \cup V$ לכל $1 \leq i \leq t$.

עבור כל כלל מהצורה הנ"ל: $(*)$ וטרמינל $a \in \Sigma$ (למשל $X_i = a$) נוסיף משתנה S_a ל- V ואת הכלל $S_a \rightarrow a$ ל- G' . נחליף את כל המופעים של a ב- $(*)$ ב- S_a . הכללים הנוותרים שאינם מצורה זו הם: $A \rightarrow X_1 X_2 \dots X_t$

כאשר $3 \leq t$ ו- $\left(\bigcup_{a \in \Sigma} \{S_a\} \right) \cup V$ לכל $1 \leq i \leq t$. עבור כל כלל כזה נוסיף משתנים חדשים (שיהיו

$$\begin{array}{l} A \rightarrow X_1 Y_1 \\ Y_1 \rightarrow X_2 Y_2 \\ \vdots \\ Y_{t-3} \rightarrow X_{t-2} Y_{t-2} \\ Y_{t-2} \rightarrow X_{t-1} X_t \end{array}$$

שונים מכלל לכלל) $Y_1, Y_2, \dots, Y_{t-2}$, ונוסיף במקומו את הכללים הבאים ל- G'' :

דוגמה:

$$\begin{array}{l} S \rightarrow Aa|a \\ A \rightarrow Ab|b|BBAa|BAa \\ B \rightarrow AS|b \end{array}$$

יהי G' דקדוק ח"ה בעל הכללים הבאים: נעביר אותו לדקדוק ח"ה בצורה הנורמלית של חומסקי ע"י האלגוריתם: (נשים לב שהדקדוק חיובי ומסתעף!)

נשאיר את הכללים: $S \rightarrow a$, $A \rightarrow b$, $B \rightarrow AS|b$ ועבור הכללים שנותרו נחליף את a ב- S_a ואת b ב- S_b כאשר S_b, S_a הם

משתנים חדשים שעבורם הוספנו את הכללים: $S_a \rightarrow a$, $S_b \rightarrow b$ ונקבל סה"כ:

$$\begin{array}{l} S \rightarrow AS_a|a \\ A \rightarrow AS_b|b|BBAS_a|BAS_a \\ B \rightarrow AS|b \\ S_a \rightarrow a \\ S_b \rightarrow b \end{array}$$

כעת במקום הכלל $A \rightarrow BBAS_a$ (הוא בעייתי כי אינו עונה על ההגדרה) נוסיף את הכללים: $A \rightarrow BY_1$, $Y_1 \rightarrow BY_2$, $Y_2 \rightarrow AS_a$

(הוספנו משתנים חדשים Y_1, Y_2) ובמקום הכלל $A \rightarrow BAS_a$, נוסיף את הכללים: $A \rightarrow BY_3$, $Y_3 \rightarrow AS_a$ נשים לב שהוספנו משתנה חדש Y_3 עם האינדקס 3 כי השמות Y_1, Y_2 תפוסים. כלומר, לכל כלל נגדיר משתנים חדשים שלא תפוסים.

לימת הניפוח לשפות חסרות הקשר (לימת בר הילל)

תהי L שפה ח"ה אזי קיים קבוע k כך שלכל $w \in L$ כך ש- $|w| \leq k$ קיימת חלוקה של w למילים $u, v, x, y, z \in \Sigma^*$ כך שמתקיים $w = uvxyz$ ומתקיים:

$$1. \quad vy \neq \varepsilon$$

$$2. \quad |vxy| \leq k$$

$$3. \quad uv^nxy^n z \in L \quad \text{לכל } n \geq 0$$

משפט: אם L שפה חסרת הקשר אזי L מקיימת את לימת הניפוח לשפות חסרות הקשר (ולא להיפך!)

משפט: משפחת השפות חסרות ההקשר סגורה תחת איחוד, שרשור, *קליני, חיתוך עם שפה רגולרית. המשפחה איננה סגורה תחת: חיתוך (בין שפות ח"ה), השלמה, הפרש.

תרגיל

הוכח ע"י לימת הניפוח לשפות ח"ה כי השפה הבאה מעל $\Sigma = \{a, b, c\}$ לא ח"ה: $L = \{w \mid \#a(w) < \#b(w) < \#c(w)\}$

פתרון: נניח בשלילה ש- L ח"ה ויהי k הקבוע המובטח מלימת הניפוח לשפות ח"ה. נביט במילה $w = a^k b^{k+1} c^{k+2}$. קל לראות ש- $w \in L$. לפי לימת הניפוח, קיימות מילים $u, v, x, y, z \in \Sigma^*$ כך ש: $w = uvxyz$ ומתקיימים תנאים 1, 2, 3 מהלימה.

מאחר ו- $|vxy| \leq k$, הרי שהמילה vxy יכולה להכיל לכל היותר 2 אותיות שונות, כלומר בכל חלוקה אפשרית vxy לא תכיל לפחות את אחת האותיות a, b או c .

נבחן את כל האפשרויות:

1. vxy לא מכילה את האות a .

2. vxy לא מכילה את האות c .

3. xy לא מכילה את האות b .

נשים לב לכך שאפשרות 3 מוכלת באפשרויות 1 ו-2, לכן מספיק לבחון אותן. (כי אנו לוקחים "חלון" בגודל k שדרכו נתבונן במילה - כמו בדוג' מההרצאה)

מקרה 1: xy לא מכילה את האות a ולכן במקרה זה $xy = b^i c^j$ כך ש- $i + j = k$. מאחר ו- $xy \neq \varepsilon$ מתקיים: $w' = uv^0 xy^0 z \notin L$, שהרי אם v או y מכילה לפחות b אחד אזי מס' ה- b 'ים ב- w' הוא לכל היותר k , ואילו מספר ה- a 'ים ב- w' הוא בדיוק k . ולכן לא יתקיים $\#a(w') < \#b(w')$. אחרת, אחת מהמילים v או y מכילות לפחות c אחד, (ושתייהן לא תכלנה b) ונקבל שלא יתקיים $\#b(w') < \#c(w')$.

מקרה 2: xy לא מכילה את האות c ולכן במקרה זה $xy = a^i b^j$ כך ש: $i + j \leq k$. מאחר ו- $xy \neq \varepsilon$ מתקיים: $w' = uv^3 xy^3 z \notin L$, שכן אם יש לפחות אות אחת (b או a) שמס' המופעים שלה גדול ב-2, ולכן מס' המופעים של a או מס' המופעים של b ב- w' הוא לפחות כמו מס' המופעים של c ב- w' .

תרגיל

נונה השפה הבאה מעל $\Sigma = \{a, b, c\}$: $L = \{X_1 C X_2 C \dots C X_k \mid k \geq 2 \text{ each } X_i \in \{a, b\}^+ \text{ and for some } i \neq j, X_i = X_j\}$ הוכיחו כי L אינה ח"ה.

פתרון: נניח בשלילה כי L ח"ה. נגדיר את השפה הבאה: $L' = L \cap L \left(\underbrace{a^* b^* c a^* b^*}_{\text{Regular Expression}} \right)$ אזי, מסגירות

תחת חיתוך עם שפה רגולרית נובע כי גם L' חסרת הקשר. נראה כי L' אינה ח"ה, נגיע לסתירה ונסיק ש- L אינה ח"ה.

מהגדרת L נובע כי: $L' = \{a^n b^m c a^n b^m \mid n + m > 0\}$, ונניח בשלילה כי L' חסרת הקשר. יהי k הקבוע המובטח מלמת הניפוח לשפות ח"ה. נבחר את המילה $w = a^k b^k c a^k b^k$, קל לראות כי $w \in L'$. לשם נוחות, נסמן $w_1 = a^k b^k$, $w_2 = a^k b^k$ כאשר $w = w_1 c w_2$. לפי לימת הניפוח קיימות מילים $u, v, x, y, z \in \Sigma^*$ כך ש: $w = uvxyz$ ומתקיימים תנאים 1, 2, 3 מהלימה. אם v או y מכילות את התו c אזי המילה $uv^2 xy^2 z \notin L$ מכילה לפחות שני מופעים של c ולכן $uv^2 xy^2 z \notin L$. נבחן את שני המקרים האפשריים כאשר גם v וגם y לא מכילות c :

1. x מכילה את התו c

2. x לא מכילה את התו c

מקרה 1: x מכילה את התו c , במקרה זה $y = a^j$, $v = b^i$ ומאחר ו- $xy \neq \varepsilon$ מתקיים $i + j > 0$ (כי ה- $|vxy|$ חייב לבוא מתוך $(a^k \underbrace{b^k c a^k}_{vxy} b^k)$). עבור המילה $uv^2 xy^2 z$ מתקיים: מס' ה- b 'ים בבלוק השמאלי גדול ממספרם בבלוק הימני או מס' ה- a 'ים בבלוק הימני גדול ממספרם בבלוק השמאלי.

מקרה 2: x לא מכילה את התו c במקרה זה $xy = a^i b^j$ כך ש: $i + j \leq k$ (כמו כן xy הוא תת מחרוזת של w_1 או של w_2 ולכן עבור המילה $uv^2 xy^2 z$ מתקיים $w_1 \neq w_2$). (מס' ה- a 'ים או מס' ה- b 'ים בבלוק הימני שונה ממספר ה- a 'ים או מס' ה- b 'ים בבלוק השמאלי, ולכן $uv^2 xy^2 z \notin L$ לא ח"ה ומכאן ש- L לא ח"ה).

שאלה למחשבה: בהרצאה ראינו כי השפה $L = \{a^i b^i c^i \mid i \geq 1\}$ לא ח"ה, כיצד נראה אם כן שהשפה $L_1 = \{a^i b^i c^i \mid i \geq 0\}$ לא ח"ה ע"י שימוש בתכונות סגירות בלבד?

תשובה: נניח כי L_1 ח"ה אז מסגירות תחת חיתוך עם שפה רגולרית, גם $L_1 \cap L(aa^*bb^*cc^*) = \{a^i b^i c^i \mid i \geq 1\}$ קיבלנו כי L ח"ה, סתירה! (ראינו כבר כי היא לא ח"ה).

שיעור 9

מעבר מדקדוק חסר הקשר חיובי לא מסתעף, G' , לדקדוק חסר הקשר חיובי ומסתעף, G'' :

נתון דקדוק חסר הקשר חיובי לא מסתעף G' , כלומר בדקדוק יש כללים מהצורה $A \rightarrow B$ (נקרא גם כלל יחידה) ואין לנו כללי ε .

נרצה למצוא דקדוק שקול G'' חיובי ומסתעף (כך שאין בו כללי יחידה וכללי ε). נראה כיצד למצוא כזה ע"י אלגוריתם שנמחיש בדוגמה הבאה:

יהי G' דקדוק חסר הקשר שכללי הגזירה שלו הם:

$$\begin{aligned} S &\rightarrow Q|b \\ Q &\rightarrow R|a \\ R &\rightarrow ba|bRa \end{aligned}$$

בדקדוק זה, אין כללי ε ולכן הוא חיובי. נבנה דקדוק ח"ה G'' שקול, שאין בו כללי ε וכללי יחידה:

בשלב הראשון: ניקח את הכללים שאינם כללי יחידה ל- G'' :

$$\begin{aligned} S &\rightarrow b \\ Q &\rightarrow a \\ R &\rightarrow ba|bRa \end{aligned}$$

בשלב השני: נמצא את כל זוגות המשתנים A, B , $A \neq B$ המקיימים: $A \xrightarrow{*}_{G'} B$ (במקרה שלנו, $S \rightarrow Q$, $Q \rightarrow R$).

לכן: $S \xrightarrow{*}_{G'} Q$ וגם $S \xrightarrow{*}_{G'} R$, מטרנזיטיביות

בשלב שלישי: איננו מכניסים לדקדוק השקול אף כלל מהצורה $A \rightarrow B$. במקום זאת:

(1) מכיוון ש: $S \xrightarrow{*}_{G'} Q$ ו- $S \xrightarrow{*}_{G'} R$, נצרף לדקדוק השקול כל כלל מהצורה: $S \rightarrow \alpha$, כאשר $Q \rightarrow \alpha$ הוא כלל ב- G' שאינו כלל יחידה או $R \rightarrow \alpha$ הוא כלל ב- G' שאינו כלל יחידה.

(2) מכיוון ש: $Q \xrightarrow{*}_{G'} R$, נצרף לדקדוק השקול כל כלל מהצורה $Q \rightarrow \alpha$ כאשר $R \rightarrow \alpha$ הוא כלל ב- G' .

שאינו כלל יחידה. (מ-1 ו-2 נובע כי במקרה שלנו, יכנסו הכללים:

מ-1: $S \rightarrow a|ba|bRa$
מ-2: $Q \rightarrow ba|bRa$

במקום הכללים:

$$\left(\begin{array}{l} S \rightarrow Q \\ Q \rightarrow R \end{array} \right)$$

סה"כ קיבלנו דקדוק חסר הקשר חיובי ומסתעף שקול, G'' , שכללי הגזירה שלו הם: $S \rightarrow b|a|ba|bRa$ (ומכיוון $Q \rightarrow a|ba|bRa$ ו)
 $R \rightarrow ba|bRa$

שאי אפשר לעבור ל- Q , אפשר להשמיטו: $\left\{ \begin{array}{l} S \rightarrow b|a|ba|bRa \\ R \rightarrow ba|bRa \end{array} \right.$

הערה: כאשר מבקשים מאיתנו לעבור מדקדוק חסר הקשר חיובי לדקדוק בצורה הנורמלית של חומסקי, תחילה יש להעביר את הדקדוק לדקדוק שקול חיובי ומסתעף, כנ"ל, ולאחר מכן יש להמשיך לפי האלגוריתם שלמדנו בשיעור הקודם עד לקבלת דקדוק שקול בצורה הנורמלית של חומסקי.

הקשר בין דקדוק בצורה הנורמלית של חומסקי ללימת הניפוח לשפות חסרות הקשר

L שפה חסרת הקשר $\Leftarrow$ אזי קיים דקדוק G' בצורה הנורמלית של חומסקי, כך ש: $L(G') = L - \{\varepsilon\}$. נסמן ב- n את מספר המשתנים ב- G' , אזי ניתן לקבוע את $k = 2^n$ להיות הקבוע המובטח מלימת הניפוח לשפות חסרות הקשר. הוכחנו בהרצאה בעזרת עובדה זו את המשפטים הבאים:¹

1. יהי G' דקדוק בצורת חומסקי בעל n משתנים, אזי $L(G')$ אינה ריקה אם קיימת מילה $w \in L(G')$ כך ש: $|w| \leq 2^n$

2. יהי G' דקדוק ח"ה בצורת חומסקי בעל n משתנים, אזי $L(G')$ לא סופית אם קיימת מילה $w \in L(G')$ כך ש: $2^n < |w| \leq 2^{n+1}$

אוטומט מחסנית:

אוטומט מחסנית הינו שייחיה $(Q, \Sigma, \Gamma, \delta, q_0, F)$ כאשר: Q, q_0, Σ, F בדומה לאס"ד / אסל"ד, Γ הוא א"ב המחסנית, $\delta: Q \times (\Sigma \cup \{\varepsilon\}) \times (\Gamma \cup \{\varepsilon\}) \rightarrow P(Q \times (\Gamma \cup \{\varepsilon\}))$, כאשר המשמעות של $(q, \beta) \in \delta(r, \sigma\alpha)$ היא, שכאשר אנו נמצאים במצב r , קוראים את התו σ ובראש המחסנית יש α אזי אפשר להוציא את התו α מהמחסנית, לכתוב במקומו β ולעבור למצב q .

קונפיגורציה: היא שלשה מעל $(Q \times \Sigma^* \times \Gamma^*)$. הקונפיגורציה (q, w, x) מתארת מצב בו נמצאים במצב q , מילת הקלט שנותר לקרוא היא w ותכולת המחסני היא x (כאשר בראש המחסנית נמצא התו הראשון ב- x).

מעבר בין קונפיגורציות: לכל $\alpha, \beta \in \Gamma \cup \{\varepsilon\}$, $\gamma \in \Gamma^*$, $\sigma \in \Sigma \cup \{\varepsilon\}$, נסמן: $(r, \sigma w, \alpha\gamma) \vdash (q, w, \beta\gamma)$ אם לכל $w \in \Sigma^*$ ולכל $q, r \in Q$ מתקיים $(q, \beta) \in \delta(r, \sigma, \alpha)$

נסמן: $(q, w, \alpha) \vdash^* (p, x, \beta)$ אם קיימת סדרת מעברים סופית בין קונפיגורציות: $(q, w, \alpha) \vdash (q_1, w_1, \alpha_1) \vdash \dots \vdash (p, x, \beta)$

השפה המתקבלת ע"י אוטומט מחסנית:

יהי אוטומט מחסנית M אזי $L(M) = \left\{ w \in \Sigma^* \mid (q_0, w, \varepsilon) \vdash_M^* (q, \varepsilon, \varepsilon), q \in F \right\}$, כלומר: כל המילים שקיימת עבורן ריצה על האוטומט שמסתיימת במצב מקבל וגם המחסנית ריקה.

¹יכול להופיע במבחן שאלות דומות

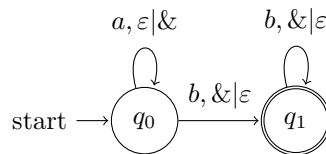
שימו לב: המודל הנ"ל הוא א־דטרמיניסטי. המודל הדטרמיניסטי לא שקול. כאשר מוזכר בקורס אוטומט מחסנית, נתכוון למודל הא־דטרמיניסטי (אלא אם צויין במפורש אחרת).

משפט: שפה L היא ח"ה אם קיים אוטומט מחסנית (א־דטרמיניסטי) M כך ש־ $L(M) = L$.

דוגמה

$$L = \{a^n b^n \mid n \geq 0\}$$

$$b, \&|\varepsilon$$



באופן כללי- סימונים במעברים בין מצב למצב:

$\underbrace{\alpha, \beta | \gamma}_{(\alpha, \beta, \gamma \neq \varepsilon)}$ - אם קראנו את התו α נוציא β מהמחסנית ונדחוף γ (לשים לב: אם אין β בראש המחסנית, המעבר לא חוקי)

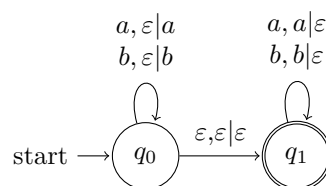
אם $\alpha = \varepsilon$ - מעבר ללא קריאת תו (תיתכן תלות במחסנית).

אם $\beta = \varepsilon$ - לא נוציא כלום מהמחסנית.

אם $\gamma = \varepsilon$ - לא נדחוף כלום למחסנית.

דוגמה

$$\Sigma = \{a, b\} \quad L = \{ww^R \mid w \in \{a, b\}^*\}$$



כאן יש שימוש בא־דטרמיניזם שבו "ניחשנו" מתי מסתיימת w ומתחילה w^R . כאן, המחסנית לא שימשה רק כמונה אלא היה גם חשוב אילו תווים הכנסנו אליה.

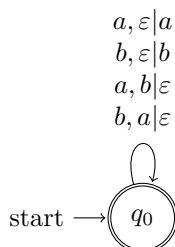
נשים לב שהמחסנית נותנת לנו להשוות מילה והיפוכה. לעומת זאת, השפה $L = \{ww \mid w \in \Sigma^*\}$ לא ח"ה (חשבו מדוע אי אפשר לבנות לה אוטומט מחסנית).

הראינו ע"י לימת הניפוח לשפות ח"ה כי השפה $L = \{a^i b^j c^I \mid i \geq 0\}$ לא ח"ה. כעת יש לנו אינטואיציה לכך, משום שאנו רואים שלא מספיקה כאן מחסנית אחת, מכיוון שלאחר ההשוואה בין ה־ a -ים ל־ b -ים, המחסנית תהיה ריקה, ולא נדע להשוות בין ה־ c -ים ל־ b -ים.

הערה: זו איננה נחשבת להוכחה!

דוגמה נוספת:

$$\Sigma = \{a, b\} \quad L = \{w \mid \#a(w) = \#b(w)\}$$



אינטואיציה:

נשתמש במחסנית כדי לזכור את ההפרש בין מס' מופעי a למס' מופעי b במילה. עם קריאת a , אפשר להכניס אותו למחסנית ובכך לזכור שהופיע ואפשר גם להוציא b במידה ומופיע בראש המחסנית כדי "לקזז" את ההפרש בין מס' המופעים. פעולות דומות נבצע בקריאת b .

שיעור 10

הוכחות נכונות

ראינו כיצד להוכיח נכונות של ביטויים רגולריים².

כיצד נוכיח נכונות של אס"ד או דקדוק?

נראה דוגמאות:

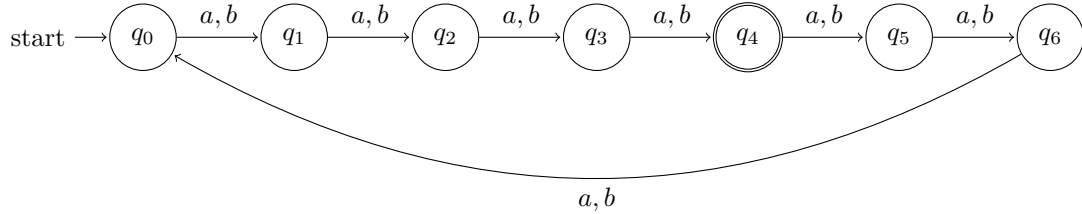
דוגמה לתרגיל עם הוכחת נכונות של אס"ד: תהי $L = \{w \mid |w| \bmod 7 = 4\}$ מעל הא"ב $\Sigma = \{a, b\}$

הוכיחו כי L רגולרית ע"י כך שתבנו אס"ד M המקבל את L והוכיחו את נכונותו.

כלומר: צריך להוכיח כי $L(M) = L$

²בתרגיל 4, כאשר נתבקשנו להוכיח שאם L רגולרית אזי L^R רגולרית.

פתרון: נבנה אוטומט M :
 $Q = \{q_i \mid i = 0, \dots, 6\}$
 $F = \{q_4\}$
 $\forall \sigma \in \Sigma, \forall q_i \in Q : \delta(q_i, \sigma) = q_{(i+1) \bmod 7}$



נוכיח כי $L(M) = L$ ע"י כך שנראה $w \in L(M) \iff w \in L$

טענת העזר: תהי מילה w מעל $\Sigma = \{a, b\}$. אם $|w| \bmod 7 = i$ אזי מתקיים $\delta^*(q_0, w) = q_i$

הוכחת טענת העזר: נוכיח באינדוקציה על $|w|$.

בסיס: $|w| = 0 \iff w = \varepsilon$ ולכן $|w| \bmod 7 = 0$, ומתקיים $\delta^*(q_0, w) = q_0$

הנחת האינדוקציה: לכל מילה מעל $\Sigma = \{a, b\}$ באורך n ($|w| = n$) , כך ש- $|w| \bmod 7 = k$ מתקיים $\delta^*(q_0, w) = q_k$

צעד האינדוקציה: תהי מילה w מעל $\Sigma = \{a, b\}$ באורך $|w| = n + 1$ ונראה כי אם $|w| \bmod 7 = i$ אז $\delta^*(q_0, w) = q_i$. נכתוב את w באופן הבא: $w = w_1\sigma$, אזי $|w_1| = n$, ולפי הנחת האינדוקציה, $\delta^*(q_0, w_1) = q_j$ וגם $|w_1| \bmod 7 = j$ עבור $0 \leq j \leq 6$ כלשהו. נתבונן באות $\sigma \in \Sigma$. עבודה מתקיים:

$$i = |w| \bmod 7 = (|w_1| + 1) \bmod 7 = ((|w_1| \bmod 7) + 1) \bmod 7 = (j + 1) \bmod 7$$

ולכן ע"פ הנחת האינדוקציה על w_1 , מתקיים: $\delta^*(q_0, w_1) = q_{(j+1) \bmod 7}$. בנוסף, ע"פ בניית M מתקיים

$$\delta^*\left(q_0, \underbrace{w_1\sigma}_w\right) = \delta(q_{(j+1) \bmod 7}, \sigma) = q_i$$

הוכחת $L(M) = L$:

(א) אם $w \in L$ אזי $|w| \bmod 7 = 4$. לפי טענת העזר, נקבל $\delta^*(q_0, w) = q_4$ ומכיוון ש- q_4 מצב מקבל, נקבל $w \in L(M)$.

(ב) אם $w \notin L$ אזי $|w| \bmod 7 = j \neq 4$. לפי טענת העזר, נקבל $\delta^*(q_0, w) = q_j \neq q_4$. מכיוון ש- q_4 מצב מקבל יחיד, נקבל כי $w \notin L(M)$.

דוגמה להוכחת נכונות דקדוק:

נתון דקדוק ח"ה $G : \varepsilon | aSb | bSa | SS$ ונתונה השפה: $L = \{w \in \{a, b\}^* \mid \#a(w) = \#b(w)\}$. הוכיחו כי $L(G) = L$.

כיוון א': $L(G) \subseteq L$: יש להוכיח $w \in L(G) \implies w \in L$. נוכיח באינדוקציה על מספר הגזירות.

טענה: לכל n ולכל $w \in L(G)$ המתקבלת ע"י סדרת גזירות באורך n מתקיים $w \in L$.

בסיס: $n = 1 \implies w = \varepsilon$ ומכאן מתקיים גם $\#a(w) = \#b(w) = 0$ ולכן $w \in L$.

הנחת אינדוקציה: נניח שלכל $k < n$ ולכל $w \in L(G)$ המתקבלת ע"י k גזירות מתקיים $w \in L$.

צעד האינדוקציה: נחלק שלב זה לשלושה מקרים: w התקבלה כך שכלל הגזירה הראשון שהופעל היה:

$$S \rightarrow aSb \quad (1)$$

$$S \rightarrow bSa \quad (2)$$

$$S \rightarrow SS \quad (3)$$

נבחן את המקרים השונים:

(1) w הוא מהצורה $w = ayb$ כך ש- $y \in L(G)$, והתקבלה ע"י סדרה של $n-1$ גזירות. לפי הנחת האינדוקציה על y , מתקיים $y \in L$ ולכן $\#a(y) = \#b(y)$. עבור המילה w מתקיים: $\#a(w) = \#a(y) + 1$ וגם $\#b(w) = \#b(y) + 1$. מכאן, מתקיים $\#a(w) = \#b(w)$ ולכן $w \in L$.

(2) ההוכחה באופן דומה למקרה א'.

(3) w היא מהצורה $w = y_1 y_2$, כך ש: $y_1, y_2 \in L(G)$. שתי מילים שהתקבלו ע"י סדרת גזירות באורך קטן מ- n . לפי הנחת האינדוקציה על y_1, y_2 מתקיים $y_1 \in L$ וגם $\#a(y_1) = \#b(y_1)$ וגם $y_2 \in L$ וגם $\#a(y_2) = \#b(y_2)$. עבור המילה w מתקיים $\#a(w) = \#a(y_1) + \#a(y_2)$ וגם $\#b(w) = \#b(y_1) + \#b(y_2)$. מכאן, מתקיים: $\#a(w) = \#b(w)$ ולכן $w \in L$.

כיוון ב': $L \subseteq L(G)$ - באינדוקציה, על אורך המילה. באתר. כמו כן יש באתר הוכחת נכונות עבור אמל"ד לשפה זו. (ראינו בשיעור קודם אמל"ד עם מצב אחד המקבל אותה).

חזרה - אוטומט מחסנית:

נתונה השפה הבאה: $L = \{x\#y \mid x, y \in \{a, b\}^* \wedge x \neq y\}$ מעל $\Sigma = \{a, b, \#\}$ בנו אוטומט מחסנית M כך ש- $L(M) = L$. הסבירו.

פתרון: נשים לב שאם $w \in L$ אז $w = x\#y$ ומתקיים אחד מהתנאים הבאים: (1) $|x| < |y|$, (2) $|x| > |y|$, (3) $|x| = |y|$. ובמקרה

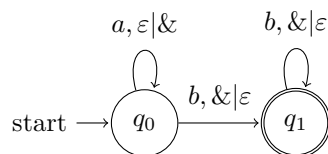
השלישי קיים אינדקס i שעבורו $x_i \neq y_i$ (כי המילים שונות).

היינו רוצים לבנות אמל"ד לכל אחד מהמקרים ולאחד את שלושתם לאמל"ד יחיד (אפשר, כי שפות ח"ה סגורות לאיחוד).

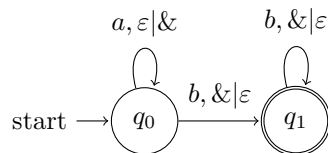
הבעיה: אי אפשר לבנות אמל"ד למקרה 3, מכיוון שנצטרך לבצע 2 סוגים שונים של ספירה, האחד לוודא כי $|x| = |y|$ והשנייה היא לספור עד לאינדקס i המקיים $x_i \neq y_i$. נשים לב כי החלק העיקרי מבחינתנו הוא למצוא אינדקס i כך ש- $x_i \neq y_i$.

לכן ננסח מחדש את מקרה 3: קיים אינדקס i שעבורו $x_i \neq y_i$ (ועדיין נדרוש $|x|, |y| \geq i$).

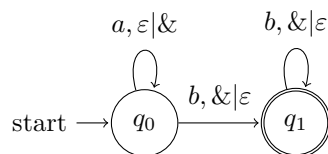
מקרה 1 - $|x| < |y|$ במקרה זה:



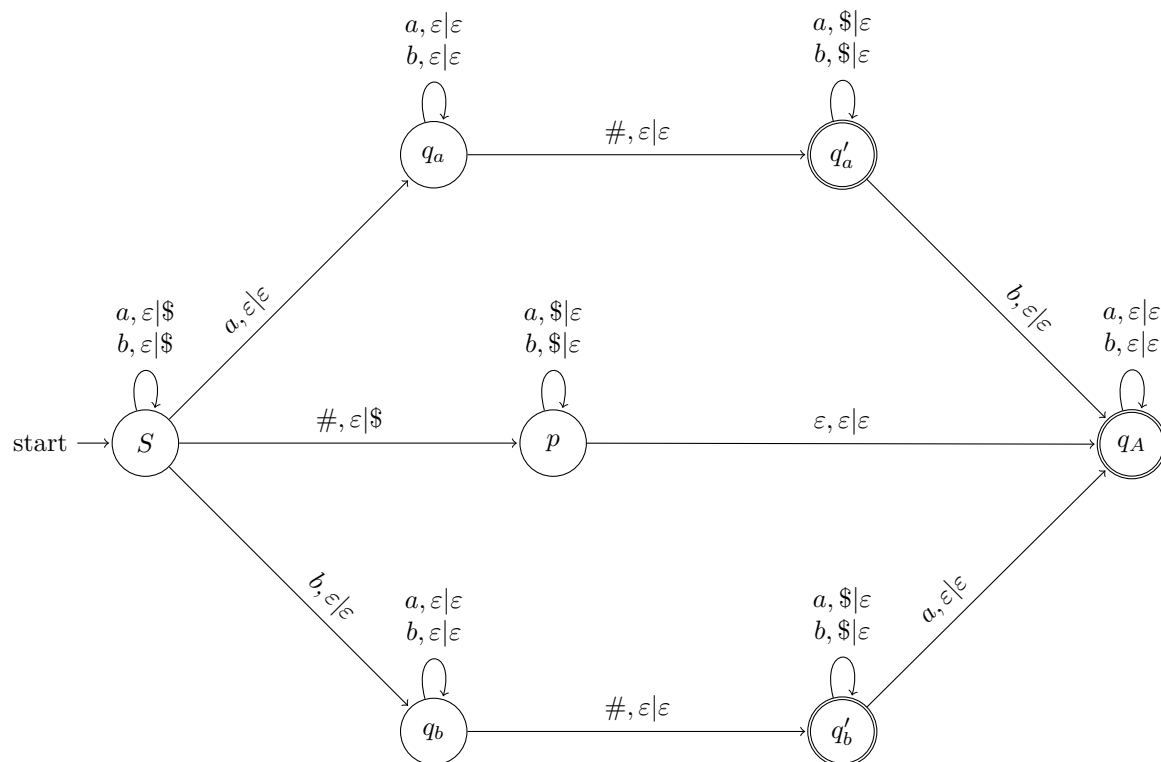
מקרה 2 - $|x| > |y|$ במקרה זה:



מקרה 2 - $|x| > |y|$ במקרה זה:



איחוד כל המקרים (וחיסכון ע"י ויתור על q ו- q') נותן:



שיעור 11

שפות תלויות הקשר Context-Sensitive languages

דקדוק תלוי הקשר

הגדרה: דקדוק תלוי הקשר הינו דקדוק שבו כל ההפקות הן מהצורה $U \rightarrow V : u, v \in (\sum \cup V)^+$ כאשר $|u| \leq |v|$.

כלומר, צד שמאל של ההפקה היא מחרוזת של טרמינלים ומשתנים (בשונה מדקדוק ח"ה שבו צד שמאל תמיד יהיה משתנה אחד). ובצד ימין של ההפקה מחרוזת של טרמינלים ומשתנים שאורכה לפחות כמו אורך המחרוזת בצד שמאל.

נשים לב: עקב דרישה זו, כללי גזירה עם ε אינם מותרים.

משפט: שפה L תלוית הקשר $\iff$ קיים דקדוק תלוי הקשר G כך ש- $L - \{\varepsilon\} = L(G)$

³ זאת, כדי להרשות ε בשפה L

אבחנה: כל שפה חסרת הקשר, היא גם תלוית הקשר.

הסבר לאבחנה: כל דקדוק חסר הקשר חיובי, הוא גם דקדוק תלוי הקשר.

דוגמה לשפה תלוית הקשר שאינה חסרת הקשר:

$L = \{a^i b^i c^i \mid i \geq 1\}$. נראה דקדוק תלוי הקשר G עבור L כך ש- $L(G) = L$:

כלל	תפקיד
$S \rightarrow aSBC$	מייצר a ויש אפשרות להמשיך
$S \rightarrow aBC$	"
$CB \rightarrow BC$	החלפת סדר b -ים לפני c -ים
$aB \rightarrow ab$	מייצר b כל עוד זה בסדר חוקי
$bB \rightarrow bb$	"
$bC \rightarrow bc$	מייצר c כל עוד זה בסדר חוקי
$cC \rightarrow cc$	"

נראה דוגמה לגזירה של $w = aabbcc$: נראה $S \xrightarrow[G]{*} w$

$S \rightarrow aSBC \rightarrow aaBCBC \rightarrow aabCBC \rightarrow aabBCC \rightarrow aabbCC \rightarrow aabbcc$

סגירויות של שפות תלויות הקשר

משפחת השפות תלויות ההקשר סגורה תחת הפעולות: איחוד, חיתוך, משלים, שרשור, *-קליני.⁴ סגירות תחת השלמה הוכח רק בסוף שנות ה-80.

Turing Machine טורינג מכונות

ע"ש אלן טורינג (Alan Turing) - מגדולי מדעני המחשב של המאה ה-20. על שמו נקרא פרס טורינג.

מכונת טורינג היא מודל מופשט לאופן פעולתו של מחשב, רעיון זה נוצר בשנת 1936 ע"י אלן טורינג עוד לפני המצאת המחשב (פיזית).

מכונת טורינג מורכבת מהרכיבים הבאים:

1. מס' סופי של מצבים
2. סרט עבודה אינסופי, בעל קצה שמאלי. הסרט מחולק לתאים, כאשר בכל תא נמצא תו מתוך Γ (נגדיר כעת).
3. ראש קורא/כותב היכול לזוז ימינה ושמאלה.

⁴חשוב לדעת למבחן!

הגדרה פורמלית: מכונת טיורינג היא שביעיה $M = (Q, \Sigma, \Gamma, \delta, q_0, q_{acc}, q_{rej})$

Q - קבוצת מצבים

Σ - א"ב מעליו אנו עובדים

q_0 - מצב התחלתי

$$\Sigma \cup \{h\} \leq \Gamma$$

q_{acc} - מצב מקבל יחיד, המכונה נעצרת כאשר היא מגיעה למצב זה.

q_{rej} - מצב דוחה יחיד, המכונה נעצרת כאשר היא מגיעה למצב זה.

עבור מכונת טיורינג דטרמיניסטית:

$$\delta : (Q \setminus \{q_{acc}, q_{rej}\} \times \Gamma) \longrightarrow Q \times \Gamma \times \{L, R, S\}$$

עבור מכונת טיורינג א-דטרמיניסטית:

$$\delta : (Q \setminus \{q_{acc}, q_{rej}\} \times \Gamma) \longrightarrow P(Q \times \Gamma \times \{L, R, S\})$$

(q', a', D) מציין שכאשר נמצאים במצב q וקוראים תו $a \in \Gamma$, ניתן לעבור למצב q' , לכתוב במקום a את a' ולהזיז את הראש הקורא/כותב לכיוון $\{L, R, S\} \in D$ כאשר R - מציין ימינה, L - שמאלה ו- S מציין השארות במקום הנוכחי.

$$L(M) = \{w \in \Sigma^* \mid \text{There exist a run of } M \text{ over the word } w \text{ in which } M \text{ stops at } q_{acc}\} : M \text{ (מ"ט) א-דטרמיניסטית (א"ד) } M$$

נשים לב: מ"ט א"ד מקבלת שפה L אם מתקיים:

- לכל $w \in L$, קיימת קיצה של M על w כך ש- M עוצרת ב- q_{acc} .
- לכל $w \notin L$, M דוחה את w או לא עוצרת על w , או נתקעת בריצה על w .

שימו לב: מ"ט אינן בהכרח עוצרות על כל קלט. כמו כן, ניתן להשתמש במ"ט על מנת לקבל פלט על גבי הסרט. (על כך יורחב בקורס המשך: "חשוביות")

מכונת טיורינג א-דטרמיניסטית חסומה לינארית⁵

מ"ט א"ד שבה הסרט מוגבל לגודל הקלט + תא ראשון משמאל לקלט המסמן התחלה ותא אחרון מימין לקלט המסמן סוף.

+	TURING	+
---	--------	---

הגדרה שקולה: הסרט מוגבל ל- $k \cdot n$ תאים כאשר n הוא גודל הקלט ו- k הוא מס' קבוע (לא תלוי בקלט) מכאן השם (לינאריות).

השקילות מתקבלת ע"י כך שניתן "לחקות" k תאים לכל תא ע"י הגדלת הא"ב של הסרט (לא נרחיב על כך כאן).

⁵נקרא גם אוטומט חסום לינארית - LBA

משפט: שפה L היא תלוית הקשר $\iff$ קיימת מ"ט א"ד חסומה לינארית M (LBA) כך ש- $L(M) = L$

שאלה פתוחה: האם מ"ט א"ד חסומה לינארית שקולה למ"ט דטרמיניסטית חסומה לינארית? בעיה בסיבוכיות: $DSPACE(O(n)) = NSPACE(O(n))$. זוהי שאלה הפתוחה משך שנים רבות.

הערה: מ"ט (כלליות - לא חסומות לינארית) דטרמיניסטיות וא-דטרמיניסטיות שקולות.

תרגיל

תנו תיאור מילולי וציירו דיאגרמה של LBA המקבל את $L = \{a^n b^n c^n \mid n \geq 0\}$ אופן פעולת המכונה על קלט w :

1. כל עוד התו הוא x המכונה מזיזה את הראש ימינה.
2. אם התו הנקרא הוא $\bar{h}$ $\Leftarrow$ מקבלת.
3. אם התו הנקרא הוא b או c $\Leftarrow$ דחה.
4. המכונה מוחקת תו a ראשון ע"י $\bar{h}$ ומזיזה את הראש ימינה.
5. כל עוד התו הוא a או x המכונה מזיזה את הראש ימינה.
- (א) אם התו הנקרא הוא c או $\bar{h}$ $\Leftarrow$ דחה.
- (ב) אם התו הנקרא הוא b , המכונה משכתבת אותו להיות x .
6. כל עוד התו הנקרא הוא b או x המכונה מזיזה את הראש ימינה.
- (א) אם התו הנקרא הוא a או $\bar{h}$ $\Leftarrow$ דחה.
- (ב) אם התו הנקרא הוא c המכונה משכתבת אותו להיות x .
7. כל עוד התו הנקרא הוא c המכונה מזיזה את הראש ימינה.
- (א) אם התו הנקרא הוא a או b $\Leftarrow$ דחה.
- (ב) אם התו הנקרא הוא $\bar{h}$, המכונה מזיזה את הראש שמאלה עד לתו הראשון שמימין לרצף הרווחים שבתחילת הסרט.
8. המכונה חוזרת לשלב 1.

משמעות המצבים: אופן פעולת המכונה הוא איטרטיבי.

קלט הלולאה: הינו המילה הכתובה מייד לאחר רצף ה- $\bar{h}$ השמאלי (אם יש) ועד תחילת תו ה- $\bar{h}$ מימין לקלט (סוף הסרט)

- q_0 - הראש מצביע על תחילת קלט הלולאה.
- q_1 - נקרא a אחד לפחות, אף לא b ואף לא c .
- q_2 - נקרא a אחד לפחות, b אחד לפחות ואף לא c .
- q_3 - נקרא a אחד לפחות, b אחד לפחות ו- c אחד לפחות.
- q_4 - נקרא a אחד לפחות, b אחד לפחות ו- c אחד לפחות, ולאחר מכן אין בקלט תווים שונים מ- c .

⁶ x הינו תו שלא נמצא בקלט, נמצא בא"ב הסרט ויסמן לנו תוים שנחקו מהקלט במהלך הריצה.

האינווריאנטה (תכונה נשמרת)

התכונות הנשמרות מאיטרציה לאיטרציה:

1. הקלט המקורי הוא מהצורה $a^*b^*c^*$ וקלט הלולאה הוא מהצורה: $a^*x^ib^*x^ic^*$
2. מספר התווים המוחקים (בין אם ע"י $\bar{h}$ או x) מרצף ה- a שווה למס' התווים המוחקים מרצף ה- b וה- c ימים.
- מגיעים למצב מקבל אם קיימת ריצה של לולאה שבסיומה מחוקים כל תווי ה- a , b ו- c והקלט הוא מהצורה $a^nb^nc^n$ עבור $0 \leq n$.

שאלה: נתונה השפה הבאה: $L = \{w\#w \mid w \in \{0,1\}^*\}$
הסבירו במילים את אלגוריתם הפעולה ש- LBA מקבל אותה.

פתרון:

הערה: אם מבקבים לתאר במילים אופן פעולה של LBA , יש לעשות זאת כמו בתרגיל קודם.

כאן בתרגיל, ביקשו לתאר אלגוריתם, ולכן ניתן לעשות זאת כך:

פתרון: אופן פעולת המכונה על קלט w :

1. המכונה בודקת אם הקלט הוא מהצורה $\{0,1\}^* \# \{0,1\}^*$ (המכונה תזיז את הראש ימינה כל עוד היא קוראת 0 או 1. אם היא הגיעה לסוף הסרט מבלי לקרוא $\#$ או לקרוא יותר מ- $\#$ אחת, דחה)
2. אם התו הנקרא הוא $\#$ המכונה בודקת כי כל הסימנים מימין לו נמחקו, אם כן נמחקו $\Leftarrow$ המכונה מקבלת.

(א) אחרת דחה.

3. יהי σ התו הנקרא במילה.

(א) המכונה זוכרת את σ וכותבת במקומו x .

(ב) המכונה מזיזה את הראש לתו הראשון ששונה מ- x מימין ל- $\#$ ובודקת אם שווה ל- σ

i. אם לא $\Leftarrow$ דחה.

ii. אחרת, המכונה כותבת במקומו x .

(ג) המכונה מזיזה את הראש שמאלה עד ה- x הראשון משמאל ל- $\#$

4. המכונה מזיזה את הראש צעד אחד ימינה

5. חוזרת למצב 2.

$$\begin{array}{cccccccccc}
0 & 1 & 1 & 0 & \# & 0 & 1 & 1 & 0 & \hbar \\
\uparrow & & & & & & & & & \\
x & 1 & 1 & 0 & \# & 0 & 1 & 1 & 0 & \hbar \\
& & & & \vdots & & & & & \\
x & x & 1 & 0 & \# & 0 & 1 & 1 & 0 & \hbar \\
& \uparrow & & & & & & & & \\
& & & & \vdots & & & & & \\
x & x & x & x & \# & x & x & x & x & \hbar \\
& & & & \uparrow & & & & & \\
x & x & x & x & \# & x & x & x & x & \hbar \\
& & & & & & & & & \uparrow
\end{array}$$