

תרגול 1

הקדמה ושפות

אוטומטים ושפות פורמליות
בר אילן תשעז 2017

עקיבא קליינרמן

אדמיניסטרציה

שעת קבלה:

יום שלישי ב14:00. בניין 109 חדר 51.

המייל שלי:

kleinea1@macs.biu.ac.il

הרכב הציון:

מבחן - 80%.

תרגילים - 20%.

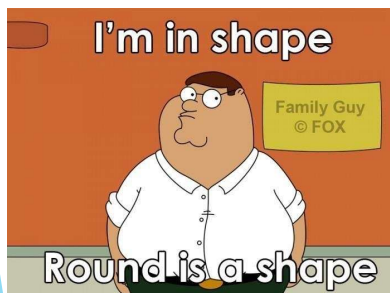
כנראה 4 תרגילים, כל אחד 5%.

הקדמה - שפות פורמליות ושפות טבעיות

▶ **התפתחות השפה:** לרוב שפות טבעיות התפתחו באופן ספונטני ולאורך זמן. שפות פורמליות עוצבו ע"י בני אדם לצורך יישומים ספציפיים.

▶ **דקדוק ותחביר (syntax):** בשפות טבעיות רבות כללי התחביר והדקדוק נוסחו בדיעבד, ואילו בשפות פורמליות נקבעו הכללים מלכתחילה.

▶ **דו-משמעות:**



✖ The method foo(int[]) is ambiguous for the type Test



המשך הקדמה

▶ **בלשנות ובלשנות חישובית.**

▶ **אוטומטים:** כלי חישובי שמקבל קלט ומגדיר האם הקלט שייך לשפה פורמלית ספציפית.

הגדרות בסיסיות

- ▶ **אות** הינה סימן כלשהו. לדוגמא λ , 8, c
- סימון מקובל: σ ו τ
- ▶ **א"ב**: אוסף סופי ולא ריק של אותיות. נסמן אותו באות Σ
- ▶ דוגמאות לא"ב: $\{1,2,\dots,8,9,0\}$ {A,B ,C,D... X,Y,Z}
- ▶ **מחרוזת/מילה**: מילה הינה סדרה סופית של אותיות מא"ב כלשהו. למשל 4532 מעל $\{0,1,2,3\dots 9\}$.
- סימון מקובל: 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" $|w|$

פעולות על מילים

- ▶ **שרשור מילים**: יהיו w_1 , w_2 מילים מעל א"ב Σ , אזי $w = w_1 w_2$ נקרא שרשור של w_1 ו w_2 .
- דוגמא: אם $w_1 = \text{bat}$ ו $w_2 = \text{man}$ אז $w_1 w_2 = \text{batman}$.
- ▶ **העלאה בחזקה**:
- יהי w_1 מילה מעל א"ב Σ אזי w_1^n הינו שרשור של w_1 n פעמים.
- דוגמא: אם $w_1 = \text{ab}$ אזי $w_1^4 = \text{abababab}$

פעולות על הא"ב

▶ עבור א"ב Σ נגדיר את Σ^n כקבוצת כל המילים מעל הא"ב Σ שאורכן n .

לדוגמא, עבור הא"ב $\Sigma = \{1,0\}$:

$$\Sigma^0 = \{\varepsilon\} \quad \Sigma^1 = \{1,0\} \quad \Sigma^2 = \{10,01,11,00\}$$

▶ Σ^* הינו קבוצת כל המילים, מכל הגדלים, מעל הא"ב Σ

נשים לב:

$$\Sigma^* = \bigcup_{n=0}^{\infty} \Sigma^n = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \dots$$

שפות

שפה הינה תת קבוצה של מילים מעל א"ב מסוים.

▶ סימון מקובל L .

▶ יכולה להיות אינסופית.

שפות לדוגמא:

▶ השפה העברית היא שפה מעל הא"ב של אותיות העברית.

▶ כל המילים מעל הא"ב $\{0,1\}$ שמכילות לפחות 3 אפסים.

$$L = \{w \mid \#_0(w) \geq 3\}$$

▶ כל המילים מעל הא"ב של אותיות העברית שנגמרות באות א.

▶ השפה הריקה $L = \emptyset$

פעולות על שפות

איחוד של השפות L_1 ו L_2 מסומן ע"י האופרטור $L_1 \cup L_2$. האיחוד מכיל את כל המילים ששייכות ל L_1 או ל L_2

$$w \in (L_1 \cup L_2) \Rightarrow w \in L_1 \vee w \in L_2$$

באופן דומה, ניתן לבצע על שפות את הפעולות חיתוך, הפרש סימטרי, חיסור, משלים וכל שאר הפעולות האפשריות על קבוצות.

$$\bar{L} = \Sigma^* \setminus L \quad \text{משלים מוגדר כך:}$$

$$L_2 \subseteq \Sigma_2^* \text{ ו } L_1 \subseteq \Sigma_1^* \quad \text{שרשור מוגדר כך:}$$

$$L_1 \cdot L_2 = \{w = w_1 \cdot w_2 \mid w_1 \in L_1, w_2 \in L_2\}$$

דוגמאות לשרשור שפות

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

$$L \cdot \emptyset = \emptyset \cdot L = \emptyset : L \quad (2) \text{ לכל שפה}$$

$$L \cdot \varepsilon = \varepsilon \cdot L = L : L \quad (3) \text{ לכל שפה}$$

לשים לב: פעולת השרשור אינה קומוטטיבית, כלומר:

$$L_1 \cdot L_2 \neq L_2 \cdot L_1$$

חזקות של שפות

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

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

כלומר, עבור n מסוים, מילה בשפה L^n היא מילה שמורכבת משרשור של n מילים מהשפה L .
זהו בעצם שרשור של שפה עם עצמה n פעמים.

לדוגמא: עבור השפה $L = \{hi, bye\}$

$$\begin{aligned} L^1 &= \{hi, bye\} \\ L^2 &= \{hihi, hibye, byehi, byebye\} \end{aligned}$$

אופרטור קליני

נגדיר את אופרטור קליני המסומן ע"י $*$:

$$L^* = \bigcup_{i=0}^{\infty} L^i$$

במילים, איחוד כל החזקות של L .

כלומר, אם $w \in L^*$ אזי ישנו n שעבורו מתקיים $w \in L^n$.

ישנו מילים $w_1, w_2, \dots, w_n \in L$ כך ש $w_1 \cdot w_2 \cdot \dots \cdot w_n = w$ ←

לדוגמא: עבור $L = \{01\}$: $L^* = \{(01)^i \mid i \geq 0\}$:

המילה הריקה מוכלת בקליני של כל שפה!

אופרטור קליני

בנוסף, נגדיר:

$$L^+ = \bigcup_{i=1}^{\infty} L^i$$

מתי זה שונה מהקליני?
בשפות שלא מכילות את המילה הריקה

תרגילים

יהיו השפות L_1 ו- L_2 : $L_1 = \{x, y, xy, yz\}$ $L_2 = \{ac, a, b, c\}$

אלו מהמילים הבאות מוכלות ב- $L_1 \cdot L_2$?

$a, cxy, cyx, x, xc, by, ayz, acz$

איך תשתנה התשובה אם נוסיף ל- L_1 את המילה הריקה? $L_1 = \{x, y, xy, yz, \varepsilon\}$

אם בשרשור של שתי שפות, שפה אחת מכילה את המילה הריקה, אז כל המילים המוכלות בשפה השנייה מוכלות גם בשרשור.

המשך תרגילים

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

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

המשך תרגילים

$$L_1(L_2 \cup L_3) = (L_1L_2) \cap (L_1L_3) \quad (2)$$

המשך תרגילים

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

Reverse

הגדרה: נגדיר פונקציה $reverse : \Sigma^* \rightarrow \Sigma^*$ באופן הבא:

$$reverse(\sigma_1 \sigma_2 \dots \sigma_n) = \begin{cases} \varepsilon & n = 0 \\ \sigma_n \sigma_{n-1} \dots \sigma_1 & n > 0 \end{cases}$$

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

תרגיל

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

תרגיל

הוכח: $(L^*)^* = L^*$

עוד 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.