פרק 5
שפות חסרות הקשר

בפרקהנוכחי נלמד על משפחה יותר כללית וחשובה של שפות פורמליות שנקראות שפות חסרות הקשר, שבאנגלית הוא Context Free Languages (CFG). רוב שפות התיכנות ושפות התוכן הן שפות חסרות הקשר, ולכן הכרת התכונות היסודיות של שפות אלה חשובה עבור קורסי המשך בשפות תיכנות, קומפיילרים, ואלגוריתמים.

5.1 דקדוק חסר הקשר

הגדרה 5.1 דקדוק חסר הקשר

הגדרה פורמלית

דקדוק חסר הקשר (CFG) G מורכב מארבעת הדברים

$$G = (V, \Sigma, R, S) \ ,$$

כאשר:

- V קבוצת **משתנים** סופית.
- Σ אלפבית סופית
- R אוסף כללי יצירה

$$A \to \alpha$$

כאשר $A \in V$ משתנה בודד בצד שמאל ו- $\alpha \in (V \cup \Sigma)^*$ מחרוזת של משתנים וטרמינילים בצד ימין.

- $S \in V$ המשתנה ההתחלתי.

יצירה של מילה על ידי דקדוק חסר קשר

1) כתבו את המשתנה ההתחלתי S.

2) מצאו משתנה וכלל אשר מתחיל אם משתנה זה, והחליפו אותו עם המחרוזת בצד ימין של הכלל.

3) חזרו על שלבים 1 ו- 2 עד שלא נשאר אף משתנים של V.

דוגמה 5.1

נתון הדקדוק חסר קשר:

$$G_1 = (\{A, B\}, \{0, 1, \#\}, R, A)$$

הקבוצת משתנים היא $V = \{A, B\}$, הקבוצת טרמינלים היא $V = \{0, 1, \#\}$, המשתנה ההתחלתי הוא $S = A$ והכללים של הדקדוק הם

$$R = \begin{cases} A \to & 0A1 \\ A \to & B \\ B \to & \# \ . \end{cases}$$

דוגמה 5.2

הדקדוק G_1 יוצר את המחרוזת $000\#111$:

$$A \xrightarrow{A\to 0A1} 0A1 \xrightarrow{A\to 0A1} 00A11 \xrightarrow{A\to 0A1} 000A111 \xrightarrow{A\to B} 00B11 \xrightarrow{B\to \#} 000\#111$$

דוגמה 5.3

נתון את הדקדוק

$$G_2 = (\{S,T,F\},\{(,),+,\times,a\},R,S)$$

כאשר הכללים הם

$$R = \begin{cases} S \to & S+T \\ S \to & T \\ T \to & T\times F \\ T \to & F \\ F \to & (S) \\ F \to & a\ . \end{cases}$$

G_2 יוצר את המילה: $a+a$:

$$S \xrightarrow{S\to S+T} S+T \xrightarrow{SA\to T} T+T \xrightarrow{T\to F} F+F \xrightarrow{F\to a} a+a$$

דוגמה 5.4 מילה המייצגת שעה בשעון

$$G = (V,\Sigma,R,S)$$

כאשר $V=\{T,H,M,L,D\}$, $\Sigma=\{0,1,2,3,4,5,6,7,8,9,:\}$, $S=T$ והכללים הם:

$$\begin{aligned} T &\to H:M\ , \\ M &\to LD\ , \\ D &\to 0|1|2|3|4|5|6|7|8|9\ , \\ L &\to 0|1|2|3|4|5\ , \\ H &\to D|10|11|12\ . \end{aligned}$$

$$T \to H:M \to H:LD \to 10:LD \to 10:4D \to 10:46\ .$$

$$T \to H:M \to H:LD \to D:LD \to 5:LD \to 5:5D \to 5:56\ .$$

דוגמה 5.5

בנו CFG היוצר את השפה $L=\left\{a^nb^n \;\middle|\; n\in\mathbb{N}\right\}$.

פתרון:

$$G = (V,\Sigma,R,S)$$

כאשר

$$
\begin{aligned}
V &= \{X\} \ , \\
\Sigma &= \{a,b\} \ , \\
S &= X \ , \\
R &= \left\{ \begin{array}{ccc} X & \to & aXb \\ X & \to & ab \\ X & \to & \varepsilon \end{array} \right\}
\end{aligned}
$$

למשל:

$$
\begin{aligned}
X &\to \varepsilon \ , \\
X &\to ab \ , \\
X \to aXb &\to aaXbb \to aabb \ .
\end{aligned}
$$

דוגמה 5.6

בנו CFG היוצר את השפה $L = \left\{a^n b^m \;\middle|\; n,m \in \mathbb{N}\right\}$.

פתרון:

$$G = (V, \Sigma, R, S)$$

כאשר:

$$
\begin{aligned}
V &= \{X, Y, S\} \ , \\
\Sigma &= \{a,b\} \ , \\
S &= S \ , \\
R &= \left\{ \begin{array}{ccc} S & \to & XY \\ X & \to & aX \\ Y & \to & bY \\ X & \to & \varepsilon \\ Y & \to & \varepsilon \end{array} \right\}
\end{aligned}
$$

למשל:

$$
\begin{aligned}
S &\to XY \to \varepsilon Y \to \varepsilon \ , \\
S &\to XY \to aXY \to a\varepsilon Y \to a\varepsilon = a \ , \\
S &\to XY \to aXY \to aXbY \to ab \ , \\
S &\to XY \to aXY \to aXbY \to aaXbY \to aab \ , \\
S &\to XY \to aXY \to aXbY \to aaXbY \to aaXbbY \to aabb \ .
\end{aligned}
$$

הגדרה 5.2 גזירות

יהי $G = (V, \Sigma, R, S)$ דקדוק חסר הקשר ויהי $(A \to \alpha) \in R$ כלל יצירה של הדקדוק.
לכל $x, y \in (V \cup \Sigma)^*$ נסמן

$$xAy \xrightarrow{G} x\alpha y \ .$$

עבור מחרוזת $x, y \in (V \cup V)^*$ נסמן

$$x \xrightarrow{G*} y$$

אם ניתן לעבור מ- x ל- y על ידי 0 או יותר גזירות של G.

הגדרה 5.3 שפת הדקדוק

יהי $G = (V, \Sigma, R, S)$.
השפה של G הינה

$$L(G) = \left\{ w \in \Sigma^* \;\middle|\; S \xrightarrow{G*} w \right\} \ .$$

אומרים כי שפה L היא **חסרת הקשר** אם קיים דקדוק חסר הקשר G כך ש- $L = L(G)$.

דוגמה 5.7

השפות הבאות הן שפות חסרות הקשר.

- $\left\{ a^n b^m \;\middle|\; m, n \in \mathbb{N} \right\}$
- $\left\{ a^n b^n \;\middle|\; n \in \mathbb{N} \right\}$
- סוגריים מאוזנים

5.2 תכנון דקדוקים חסרי הקשר

דוגמה 5.8

מצאו דקדוק שיוצר את השפת המילים

$$L = \left\{ a^i b^j a^k b^{i+j+k} \;\middle|\; i, j, k \geqslant 0 \right\} \ .$$

פתרון:
ראשית נציג את השפה בצורה פשוטו יותר:

$$L = \left\{ a^i b^j a^k b^k b^j b^i \;\middle|\; i, j, k \geqslant 0 \right\} \ .$$

$$G = (V, \Sigma, R, S)$$

כאשר:

$$\begin{aligned} V &= \{A, B, S\} \ , \\ \Sigma &= \{a, b\} \ , \\ S &= S \ , \\ R &= \left\{ \begin{array}{ccc} S & \to & aSb|A \\ A & \to & bAb|B \\ B & \to & aBa|\varepsilon \end{array} \right\} \end{aligned}$$

למשל יצירת המילה $a^2b^1a^1b^4$:

$$S \to aSb \to aaSbb \to aaAbb \to aabAbbb \to aabAbbb \to aabBbbb \to aabaBbbbb \to aababbbb$$

דוגמה 5.9

מצאו דקדוק שיוצר את השפת המילים

$$L = \left\{ w \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, ", "\}^* \;\middle|\; w \text{ לא מתחילה בספרה } 0 \right\} \ .$$

כלומר הספרות ובנוסף הפסיק. לדוגמה:

$$14,057,121 \in L \ .$$

הבהרה: אפסים מקדימים אינם חוקיים ולכן $01,234 \notin L$.

פתרון:

$$G = (V, \Sigma, R, S)$$

כאשר:

$$\begin{aligned}
V =& \{S, X, Y, Z\} \ , \\
\Sigma =& \{0,1,2,3,4,5,6,7,8,9,",\} \ , \\
S =& S \ , \\
R =& \left\{ \begin{array}{ccc} S & \rightarrow & 0|XY \\ Y & \rightarrow & ZY|\varepsilon \\ X & \rightarrow & 1|2|3|4|5|6|7|8|9 \\ Z & \rightarrow & 0|1|2|3|4|5|6|7|8|9|, \end{array} \right\}
\end{aligned}$$

למשל יצירת המילה 1024:

$$S \rightarrow XY \rightarrow XZY \rightarrow XZZY \rightarrow XZZZY \rightarrow XZZZ \rightarrow 1024 \ .$$

יצירת המילה $4,096$:

$$S \rightarrow XY \rightarrow XZY \rightarrow XZZY \rightarrow XZZZY \rightarrow XZZZ \rightarrow XZZZZ \rightarrow 4,096 \ .$$

דוגמה 5.10

בנו CFG לשפה $L = \left\{a^n b^m \ \middle| \ n \neq m\right\}$.

פתרון:

$$G = (V, \Sigma, R, S)$$

כאשר:

$$\begin{aligned}
V =& \{S, A, B\} \ , \\
\Sigma =& \{a, b\} \ , \\
S =& S \ , \\
R =& \left\{ \begin{array}{ccc} S & \rightarrow & aSb|aA|bB \\ A & \rightarrow & aA|\varepsilon \\ B & \rightarrow & bB|\varepsilon \end{array} \right\}
\end{aligned}$$

למשל יצירת המילה aab:

$$S \rightarrow aSb \rightarrow aaAb \rightarrow aab \ .$$

יצירת המילה $aabbb$:

$$S \rightarrow aSb \rightarrow aaSbb \rightarrow aabBbb \rightarrow aabbb \ .$$

דוגמה 5.11

בנו CFG לשפת כל המילים מעל הא"ב $\{a,b\}$ המכילות את המחרוזת $abba$.

פתרון:

$$G=(V,\Sigma,R,S)$$

כאשר:

$$\begin{aligned}
V &= \{S,A\} \ , \\
\Sigma &= \{a,b\} \ , \\
S &= S \ , \\
R &= \left\{ \begin{array}{ccc} S & \rightarrow & AabbaA \\ A & \rightarrow & aA|bA|\varepsilon \end{array} \right\}
\end{aligned}$$

למשל יצירת המילה $babbaa$:

$$S \rightarrow AabbaA \rightarrow babbaa \ .$$

דוגמה 5.12

בנו CFG לשפה

$$L=\left\{w\in\{a,b\}^* \ \middle| \#a_w=2\right\} \ .$$

פתרון:

$$G=(V,\Sigma,R,S)$$

כאשר:

$$\begin{aligned}
V &= \{S,A\} \ , \\
\Sigma &= \{a,b\} \ , \\
S &= S \ , \\
R &= \left\{ \begin{array}{ccc} S & \rightarrow & AaAaA \\ A & \rightarrow & bA|\varepsilon \end{array} \right\}
\end{aligned}$$

דוגמה 5.13

בנו CFG לשפה

$$L=\left\{w\in\{a,b\}^* \ \middle| \#a_w\leqslant 2\right\} \ .$$

פתרון:

$$G=(V,\Sigma,R,S)$$

כאשר:

$$
\begin{aligned}
V =& \{S, A\} \ , \\
\Sigma =& \{a, b\} \ , \\
S =& S \ , \\
R =& \left\{ \begin{array}{ccc} S & \rightarrow & A|AaA|AaAaA \\ A & \rightarrow & bA|\varepsilon \end{array} \right\}
\end{aligned}
$$

- יצירת המילה $babb$:

$$S \rightarrow AaA \rightarrow bAaA \rightarrow bAabAbA \rightarrow babb \ .$$

דוגמה 5.14

בנו CFG לשפה

$$L = \left\{ w \in \{a, b\}^* \ \middle| w \neq w^r \right\} \ .$$

פתרון:

$$G = (V, \Sigma, R, S)$$

כאשר:

$$
\begin{aligned}
V =& \{S, A\} \ , \\
\Sigma =& \{a, b\} \ , \\
S =& S \ , \\
R =& \left\{ \begin{array}{ccc} S & \rightarrow & aSa|bSb|bAa|aAb \\ A & \rightarrow & aA|bA|\varepsilon \end{array} \right\}
\end{aligned}
$$

הסבר: כדי לסיים את הגזירה של מילה חייבים לגזור משתנה A, וזה מתאפשר רק כאשר הסימטריות של המילה מופרת $(aAb|bAa)$.

- יצירת המילה baa:

$$S \rightarrow bAa \rightarrow baAa \rightarrow baa \ .$$

- למשל יצירת המילה abb:

$$S \rightarrow aAb \rightarrow abAb \rightarrow abb \ .$$

דוגמה 5.15

בנו CFG לשפה

$$L = \left\{ w \in \{a, b\}^* \ \middle| \#a_w > \#b_w \right\} \ .$$

פתרון:

$$G = (V, \Sigma, R, S)$$

כאשר:

$$
\begin{aligned}
V &= \{S, X, Y, Z\} \ , \\
\Sigma &= \{a, b\} \ , \\
S &= S \ , \\
R &= \left\{ \begin{array}{ccc} S & \to & XaX \\ X & \to & XYZ|XZY|\varepsilon \\ Y & \to & aY|\varepsilon \\ Z & \to & ZZ|aZb|bZa|\varepsilon \end{array} \right\}
\end{aligned}
$$

הסבר:

- נשים לב שכל w ניתנת לחלוקה $w = xay$ כך שב x, y מספר ה- a לפחות כמספר ה- b. מחרוזות אלו נוצרות ע"י המשתנה X.
- כל מחרוזת שבה מספר ה- a לפחות כמספר ה- b ניתנת לחלוקה לחלקים כך שכל חלק יהיה באחת משתי צורות הבאות:
  1) מספר שווה של אותיות a ו- b,
  2) רצף אותיות a.
- המשתנה Z יוצר מחרוזות מסוג 1. המשתנה Y יוצר מחרוזות מסוג 2.
- יצירת המילה aab:

$$S \to XaX \to \varepsilon aX \to aXYZ \to a\varepsilon Z \to aaZb \to aab \ .$$

משפט 5.1 למת הניפוח לשפות חסרות הקשר

תהי L שפה חסרת הקשר.
קיים מספר טבעי p (קבוע למת הניפוח) כך שלכל $w \in L$ $(|w| \geqslant p)$ קיים פירוק $w = tuxyz$ כך שהתנאים הבאים מתקיימים:

1) $|uy| > 0$,
2) $|uxy| \leqslant p$,
3) לכל i טבעי: $tu^ixy^iz \in L$.

הוכחה:

■

דוגמה 5.16

תהי $L = \left\{a^nb^nc^n \;\middle|\; n \in \mathbb{N}^+\right\}$. הוכיחו כי $L \notin CFG$.

פתרון:

- נניח בשלילה כי $L \in CFG$.
- נבחור $w = a^pb^pc^p \in L$ כאשר p קבוע למת הניפוח.

- לכל פירוק $w = tuxyz$ המקיים התנאי למת הניפוח, בפרט $|uy| > 0$ וגם $|uxy| \leqslant p$, המחרוזת uxy כוללת שתי אותיות שונות לכל היותר ולכן 5 מקרים:

$$\begin{array}{llll}
t = a^j \ , & uxy = a^k \ , & z = a^{p-j-k}b^pc^p \ , & k > 0 \ \wedge \ (j+k \leqslant p) \ , \\
t = a^pb^j \ , & uxy = b^k \ , & z = b^{p-j-k}c^p \ , & k > 0 \ \wedge \ (j+k \leqslant p) \ , \\
t = a^pb^pc^j \ , & uxy = c^k \ , & z = c^{p-j-k} \ , & k > 0 \wedge (j+k \leqslant p) \ , \\
t = a^{p-j} \ , & uxy = a^jb^k \ , & z = b^{p-k}c^p \ , & j,k \leqslant p \wedge \ (j > 0 \vee k > 0) \ , \\
t = a^pb^{p-j} \ , & uxy = b^jc^k \ , & z = c^{p-k} \ , & j,k \leqslant p \wedge \ (j > 0 \vee k > 0) \ .
\end{array}$$

- בכל מקרה, עבור אורך הניפוח $i = 2$ נקבל המילה $w' = tu^2xy^2z$ כל ש:

$$\begin{array}{ll}
w' = a^nb^pc^p & (n > p) \\
w' = a^pb^nc^p & (n > p) \\
w' = a^pb^pc^n & (n > p) \\
\Big((w' = a^nb^pc^p) \quad \vee \quad (w' = a^pb^nc^p) \quad \vee \quad (w' = a^nb^mc^p) \Big) & (n, m > p) \\
\Big((w' = a^pb^nc^p) \quad \vee \quad (w' = a^pb^pc^n) \quad \vee \quad (w' = a^pb^nc^m) \Big) & (n, m > p) \ .
\end{array}$$

$\Leftarrow$ $tu^2xy^2z \notin L$

$\Leftarrow$ סתירה לכך שעל פי למת הניפוח: $tu^2xy^2z \in L$.

דוגמה 5.17

1) הראו שהשפה $L_1 = \{a^ib^jc^jd^i \mid i,j \geqslant 0\}$ היא שפה חסרת הקשר על ידי הגדרת דקדוק חסר הקשר מתאים.

2) הראו שהשפה $L_2 = \{a^ib^jc^id^j \mid i,j \geqslant 0\}$ אינה שפה חסרת הקשר על ידי למת אניפוח לשפות חסרות מתאים.

פתרון:

סעיף 1)

$$G = (V, \Sigma, S, R)$$

- $V = \{S, X\}$
- $\Sigma = \{a, b\}$
- $R = \left\{ \begin{array}{lll} S & \rightarrow & aSd|X \ , \\ X & \rightarrow & bXc|\varepsilon \ . \end{array} \right\}$

סעיף 2)

- נניח בשלילה L_2 בפה חסרת הקשר.
- תהי $w = a^pb^pc^pd^p$ כאשר p קבוע למת הניפוח.
- כל פירוק שמקיים התנאי למת הניפוח הוא בהכרח מהצורה $w = tuxyz$ כאשר $|uxy| \leqslant p$ וגם $|uy| > 0$.
- $0 < |uy| \leqslant p$ ולכן המחרוזת uy מכילה רצף של אות אחת או רצפי שתי אותיות.
  בכל מקרה בתוך uy לא יופיע שתי אותיות לא עוקבות.

- תהי $w' = tu^2xy^2z$. יש 4 מקרים:

1)
* uy מכיל אותיות a (לפחות אתת) או אותיות a (לפחות אתת) וגם אותיות b.
* לא ייתכן ש- uy מכיל אותיות c ו- d.
* במילה $w' = tu^2xy^2z$ מספר האותיות a גדל (כמספר אותיות ה- a במחרוזת uy) בעוד מספר האותיות c נשאר כפי שהוא.
* לכן:

$$\#a_{w'} = p + \#a_{uy} \geqslant p + 1 = \#c_{w'}$$

* לכן $w' \notin L$.

2)
* uy מכיל אותיות b (לפחות אתת) או אותיות b (לפחות אתת) וגם אותיות c.
* uy לא מכיל אותיות a ו- d.
* במילה $w' = tu^2xy^2z$ מספר האותיות b גדל (כמספר אותיות ה- b במחרוזת uy) בעוד מספר האותיות d נשאר כפי שהוא.
* לכן:

$$\#b_{w'} = p + \#b_{uy} \geqslant p + 1 = \#d_{w'}$$

* לכן $w' \notin L$.

3)
* uy מכיל אותיות c (לפחות אתת) או אותיות c (לפחות אתת) וגם אותיות d.
* uy לא מכיל אותיות a ו- b.
* במילה $w' = tu^2xy^2z$ מספר האותיות c גדל (כמספר אותיות ה- c במחרוזת uy) בעוד מספר האותיות a נשאר כפי שהוא.
* לכן:

$$\#c_{w'} = p + \#c_{uy} \geqslant p + 1 = \#a_{w'}$$

* לכן $w' \notin L$.

4)
* uy מכיל רק אותיות d (לפחות אתת).
* במילה $w' = tu^2xy^2z$ מספר האותיות d גדל (כמספר אותיות ה- d במחרוזת uy) בעוד מספר האותיות b נשאר כפי שהוא.
* לכן:

$$\#d_{w'} = p + \#d_{uy} \geqslant p + 1 = \#b_{w'}$$

* לכן $w' \notin L$.

- בכל מקרה המילה $w' \notin L$ בסתירה לכך שעל פי למת הניפוח המילה $tu^2xy^2z \in L$.