אוטומטים ושפות פורמליות – 236353
סמסטר אביב 2016

קובץ שאלות ופתרונות של שאלות ממבחנים מנושאים שונים

קובץ ונערך ע״י אורן אשכנזי ומיכל הורוביץ

1. תכונות סגור ודקדוקים רגולריים.

עבור שפות L_1, L_2 מעל א״ב Σ נגדיר את השפה $L = \{uv : u \in L_1, v \in L_2, |u| = |v|\}$.

הוכיחו/הפריכו את הטענות הבאות באמצעות תכונות סגור בלבד:

א. אם L_1, L_2 רגולריות אז L רגולרית.

ב. אם L_1 רגולרית ו-L_2 ח״ה אז L ח״ה.

בסעיפים הבאים הוכיחו כי אם L_1, L_2 רגולריות אז L ח״ה.

ג. באמצעות תכונות סגור.

ד. ע״י בניית דקדוק ח״ה מתאים.

הערה: מספיק לנמק את נכונותו, ואין צורך להוכיח באופן מלא את שפת הדקדוק.

פתרון:

א. לא נכון!

דוגמא נגדית: $L_1 = \{a\}^*, L_2 = \{b\}^*$ מעל א״ב $\{a, b\}$.

- השפות L_1, L_2 רגולריות מכיוון שהשפות $\{a\}, \{b\}$ סופיות ולכן רגולריות, ומסגירות השפות הרגולריות לאיטרציה. כמו-כן מתקיים:

$$L = \{uv : u \in \{a\}^*, v \in \{b\}^*, |u| = |v|\} = \{a^i b^j : i, j \geq 0, |a^i| = |b^j|\} = \{a^n b^n : n \geq 0\}$$

- ידוע שהשפה $\{a^n b^n : n \geq 0\}$ אינה רגולרית, כלומר L אינה רגולרית, בניגוד לטענה.

ב. לא נכון!

דוגמא נגדית: $L_1 = \{c\}^*, L_2 = \{a^n b^n : n \geq 0\}$ מעל א״ב $\{a, b, c\}$.

- השפה L_1 רגולרית מכיוון ש-$\{c\}$ שפה סופית ולכן רגולרית, ומסגירות השפות הרגולריות לאיטרציה. כמו-כן ידוע שהשפה L_2 היא ח״ה.
- מתקיים: $L = \{uv : u \in \{c\}^*, v \in \{a^n b^n : n \geq 0\}, |u| = |v|\} = \{c^{2n} a^n b^n : n \geq 0\}$.
- נניח בשלילה ש-L שפה חסרת הקשר.
- נגדיר הומומורפיזם $h : \{a, b, c\} \to \{a, b, c\}^*$ באופן הבא: $h(a) = cc, h(b) = a, h(c) = b$. נשים לב שמתקיים: $h^{-1}(L) = \{w : \exists n \geq 0, h(w) = c^{2n} a^n b^n\} = \{a^n b^n c^n | n \geq 0\}$. שפה זו היא חסרת הקשר מסגירות השפות חסרות ההקשר להומומורפיזם הפוך. אך זה בסתירה לכך שידוע שהשפה $\{a^n b^n c^n | n \geq 0\}$ אינה חסרת הקשר! לכן קיבלנו ש-L אינה ח״ה, בניגוד לטענה.

ג. נכון!

- נסמן $\Sigma' = \{\sigma' : \sigma \in \Sigma\}$. כמו-כן נסמן ב-$w'$ את המילה w בתוספת תיוג על כל אות במילה.

- נגדיר הומומורפיזם $h_1: \Sigma \to \Sigma'^*$ באופן הבא: $\forall \sigma \in \Sigma: h_1(\sigma) = \sigma'$. נשים לב שלכל מילה w מתקיים $h_1(w) = w'$. נגדיר: $L_2' = h_1(L_2) = \{w' : w \in L_2\}$. שפה זו היא רגולרית מסגירות השפות הרגולריות להומומורפיזם.
- נסמן $L_{ab} = \{a^n b^n : n \geq 0\}$, ידוע ששפה זו היא חסרת הקשר.
- נגדיר הצבה $f: \{a, b\} \to 2^{(\Sigma \cup \Sigma')^*}$, באופן הבא: $f(a) = \Sigma, f(b) = \Sigma'$. נשים לב כי $f(a), f(b)$ שתיהן שפות סופיות, לכן הן רגולריות ולכן גם חסרות הקשר, ולכן ההצבה f היא הצבה ח״ה. נגדיר: $L_3 = f(L_{ab}) = \{f(a^n b^n) : n \geq 0\} = \{uv' : u, v \in \Sigma^*, |u| = |v|\}$. שפה זו היא חסרת הקשר מסגירות השפות חסרות ההקשר להצבה ח״ה.
- נגדיר: $L_4 = L_1 L_2' = \{uv' : u \in L_1, v' \in L_2'\} = \{uv' : u \in L_1, v \in L_2\}$. שפה זו היא רגולרית מסגירות השפות הרגולריות לשרשור סופי.
- נגדיר $L_5 = L_3 \cap L_4 = \{uv' : u \in L_1, v \in L_2, |u| = |v|\}$. שפה זו היא חסרת-הקשר מסגירות השפות חסרות ההקשר לחיתוך עם שפות רגולריות.
- נגדיר הומומורפיזם $h_2: \Sigma \cup \Sigma' \to \Sigma^*$ באופן הבא: $\forall \sigma \in \Sigma: h_2(\sigma) = h_2(\sigma') = \sigma$. נשים לב ש: $h_2(L_5) = \{h_2(uv') : u \in L_1, v \in L_2, |u| = |v|\} = \{uv : u \in L_1, v \in L_2, |u| = |v|\} = L$.
לכן קיבלנו שהשפה L היא חסרת הקשר מסגירות השפות חסרות ההקשר להומומורפיזם.

ד. <u>רעיון הבנייה</u>: ניקח דקדוקים רגולריים עבור השפות L_1, L_2 ונשלב את הכללים שלהם לבניית דקדוק חסר הקשר. הכללים של הדקדוקים הנ״ל גוזרים את המילים אות אחר אות. אז כדי לגזור $uv \in L$, נגזור את u ואת v בו-זמנית, אות אחר אות, וכך נוכל לדאוג שהן מאותו אורך. כדי לשלוט על שתי הגזירות בו-זמנית, צריך משתנה דקדוקי משותף שמבצע את שתיהן, והוא חייב להיות ממוקם באמצע בין u ל-v לכל אורך הגזירה. לכן נגזור את u משמאל לימין ואת v מימין לשמאל, ע״י כך שניקח דקדוקים לינארי ימני ולינארי שמאלי בהתאמה.

<u>בניה</u>: נניח שמתקיים $\epsilon \notin L_1$, $\epsilon \notin L_2$ (נתייחס בהמשך למקרים שזה לא המצב). ניקח דקדוק לינארי ימני $G_1 = (V_1, \Sigma, P_1, S_1)$ כך ש-$L(G_1) = L_1$ ודקדוק לינארי שמאלי $G_2 = (V_2, \Sigma, P_2, S_2)$ כך ש-$L(G_2) = L_2$ (שני הדקדוקים קיימים כי השפות רגולריות). כמו-כן נוודא ששני הדקדוקים ללא כללי ϵ (זה אפשרי כי ראינו שתמיד קיים דקדוק רגולרי שכלל ה-ϵ היחיד בו הוא $S \to \epsilon$, וכאן אין כזה כי הנחנו ש-ϵ לא בשפות). כעת נבנה את הדקדוק הבא:

$$G = (V_1 \times V_2, \Sigma, P, (S_1, S_2))$$
$$P = \{(X_1, X_2) \to \sigma_1 (Y_1, Y_2) \sigma_2 \mid X_1 \to \sigma_1 Y_1 \in P_1,\ X_2 \to Y_2 \sigma_2 \in P_2\}$$
$$\cup \{(X_1, X_2) \to \sigma_1 \sigma_2 \mid X_1 \to \sigma_1 \in P_1,\ X_2 \to \sigma_2 \in P_2\}$$

<u>הסבר לבניה</u>: המשתנים הם זוגות של משתנים מ-V_1, V_2, כך אנחנו יכולים לשלוט על שתי הגזירות בו זמנית. בכל פעם נגזור אות אחת של u לפי כלל מ-P_1 ואות אחת של v לפי כלל מ-P_2 ונעדכן את המשתנה המשותף בהתאם. אם שני המשתנים הנוכחיים מאפשרים גזירה טרמינלית אז נאפשר גם גזירה טרמינלית מהמשתנה המשותף.
לפיכך נקבל:

$w \in L \Leftrightarrow$ קיימים $u \in L_1, v \in L_2$ כך ש: $w = uv$ ו-$|u| = |v|$

$\Leftrightarrow$ קיימים $u, v \in \Sigma^*$ כך ש: $S_1 \Rightarrow_{G_1}^k u$, $S_2 \Rightarrow_{G_2}^k v$ ו- $k = |u| = |v|$, $w = uv$

$$\xLeftrightarrow{by\ the\ above\ explanation} (S_1, S_2) \Rightarrow_G^k w \iff w \in L(G)$$

<u>טיפול ב-ϵ</u>: מפתה לטפל בכללי ϵ ע״י הוספת: $\{(X_1,X_2)\rightarrow\epsilon \mid X_1\rightarrow\epsilon\in P_1, X_2\rightarrow\epsilon\in P_2\}$ לכללי הגזירה בבניה הנ״ל. אבל נשים לב שזה יוצר מקרה בעייתי. למשל יתכן שב-G_1 הכללים הם $S_1\rightarrow aS_1|\epsilon$, ואז קיימת הגזירה: $S_1\Rightarrow aS_1\Rightarrow a$, וב-$G_2$ הכלל היחיד הוא $S_2\rightarrow b$, ואז קיימת הגזירה: $S_2\Rightarrow b$, אבל ב-G לא נקבל גזירה למילה ab כי מדובר בגזירות באורך שונה. לכן לצורך הפשטות לקחנו מראש דקדוקים ללא כללי ϵ, ואז ברור שאורך הגזירה שווה לאורך המילה והכל מסתדר יפה. במידה וכן מתקיים $\epsilon\in L_1$ או $\epsilon\in L_2$, פשוט נבצע את הבניה הנ״ל עבור $L_1\setminus\{\epsilon\}$ ו-$L_2\setminus\{\epsilon\}$, ונקבל דקדוק ח״ה עבור $L\setminus\{\epsilon\}$, ממנו ניתן בקלות לבנות דקדוק ח״ה עבור L (ע״י בנית דקדוק לאיחוד שפות שנלמדה בכיתה).

2. תכונות סגור ולמת הניפוח לשפות רגולריות.

תהי $L=\{a^ib^j \mid i\neq j\in\mathbb{N}\}$.

א. הוכיחו/הפריכו ע״י תכונות סגור בלבד: L רגולרית.

ב. הוכיחו/הפריכו: L מקיימת את למת הניפוח לשפות רגולריות.

<u>פתרון</u>:

א. L אינה רגולרית!

- נניח בשלילה שהשפה L היא רגולרית.
- השפה $\{a\}^*\{b\}^*$ רגולרית מכיוון ש-$\{a\},\{b\}$ סופיות ולכן רגולריות, ומסגירות השפות הרגולריות לשרשור סופי ולאיטרציה.
- השפה $\{a\}^*\{b\}^*-L$ רגולרית מסגירות השפות הרגולריות לחיסור. נשים לב כי:

$$\{a\}^*\{b\}^*-L=\{w \mid w\in\{a\}^*\{b\}^*, w\notin L\}=\{a^ib^j \mid i,j\geq 0, a^ib^j\notin L\}=\{a^nb^n \mid n\geq 0\}$$

- ידוע שהשפה $\{a^nb^n \mid n\geq 0\}$ אינה רגולרית ולכן קיבלנו סתירה. לכן L אינה רגולרית.

ב. השפה L אינה מקיימת את למת הניפוח.

- נניח בשלילה שהיא כן מקיימת, ויהי n הקבוע המובטח מהלמה.
- נביט במילה $z=a^nb^{n+n!}$, עבורה מתקיים $|z|\geq n$ וגם $z\in L$ (כי $n!>0$ תמיד).
- לפי הלמה קיים פירוק $z=uvw$ המקיים את תנאיה. מכיוון ש-$|uv|\leq n$ אז בהכרח הפירוק מהצורה: $u=a^s$, $v=a^t$, $w=a^{n-s-t}b^{n+n!}$. מכיוון ש-$|v|\geq 1$ מתקיים $t\geq 1$. כמו-כן מתקיים $|v|\leq|uv|\leq n$, לכן גם $t\leq n$.
- לפי הלמה $z_i\in L$ לכל $i\in\mathbb{N}$, בפרט עבור $i=\frac{n!}{t}+1$ (זה מספר שלם מכיוון ש-$1\leq t\leq n$). אבל:

$$z_i=uv^iw=a^sa^{ti}a^{n-s-t}b^{n+n!}=a^sa^{n!+t}a^{n-s-t}b^{n!+n}=a^{n!+n}b^{n!+n}\notin L$$

- זו סתירה. לכן קיבלנו שהשפה L אינה מקיימת את למת הניפוח.

3. למת הניפוח לשפות רגולריות או בנית אוטומט.

יהי Σ א״ב כך ש-$|\Sigma|\geq 2$. עבור $u\in\Sigma^*$, נסמן $perm(u)=\{v\in\Sigma^*: \forall\sigma\in\Sigma, \#_\sigma(u)=\#_\sigma(v)\}$, כלומר $perm(u)$ היא קבוצת כל המילים המתקבלות מ-u ע״י פרמוטציה של האותיות.

נגדיר את השפה: $L=\{w\in\Sigma^*: \exists u,u'\in\Sigma^*, w=uu', u'\in perm(u)\}$.

<u>הוכיחו/הפריכו</u>: L רגולרית.

<u>הנחיה</u>: כדי להוכיח, עליכם <u>לבנות אוטומט סופי</u>, <u>ולנמק</u> את נכונות הבנייה, אך <u>אין</u> צורך בהוכחת נכונות פורמלית. כדי <u>להפריך</u>, יש להשתמש <u>בלמת הניפוח</u> לשפות רגולריות.

פתרון:
<u>השפה אינה רגולרית!</u>

- נניח בה"כ ש-$\{a,b\} \subseteq \Sigma$ (נתון ש-$|\Sigma| \geq 2$, סימני האותיות אינם מגבילים כלליות).
- נניח בשלילה ש-L רגולרית, לכן היא מקיימת את למת ניפוח. יהי n הקבוע המובטח מהלמה.
- נביט במילה $z = a^n bba^n$. נשים לב כי $|z| \geq n$ וכן $z \in L$ (כי מתקיים $a^n b \in perm(ba^n)$).
- לפי הלמה קיים פירוק $z = uvw$ המקיים את תנאיה. מכיוון ש-$|uv| \leq n$ הפירוק הוא בהכרח מהצורה $u = a^s$, $v = a^t$, $w = a^{n-s-t}bba^n$. כמו כן $|v| \geq 1$ לכן $t \geq 1$.
- לפי הלמה, לכל i מתקיים $z_i \in L$, בפרט: $z_0 = uv^0w = a^s a^{n-s-t} bba^n = a^{n-t}bba^n \in L$.
- לפי הגדרת L, קיימת $u \in \Sigma^*$ כך ש: $z_0 = uu'$ וגם: $u' \in perm(u)$.
כלומר $\#_b(u) = \#_b(u') = 1$ (כי $\#_b(z_0) = 2$). אך הפירוק היחיד של z_0 שמקיים את זה הוא: $u = a^{n-t}b$, $u' = ba^n$. אבל $\#_a(u) = n - t < n = \#_a(u')$ בסתירה לכך ש-$u' \in perm(u)$.

$\Leftarrow$ L אינה רגולרית.

4. הוכחה מההרצאה. הוכחת נכונות שפה של אוטומט.
הוכיחו כי השפות הרגולריות סגורות תחת הפעולה של הומומורפיזם הפוך.

פתרון:
<u>אינטואיציה</u>: אנחנו רוצים לבנות אוטומט שמקבל את השפה $h^{-1}(L)$. כלומר אוטומט שמקבל מילה w אמ"מ $h(w)$ היא מילה ב-L. כדי להכריע האם $h(w)$ היא מילה ב-L או לא, נריץ סימולציה של אס"ד שמקבל את L. יש לשים לב שהקלט לאוטומט שאנחנו בונים הוא המילה w עצמה, ולכן על כל אות σ שמופיעה בקלט, נריץ את הסימולציה "בבת אחת" על כל האותיות ב-$h(\sigma)$. לצורך כך נסתפק באותם מצבים של האוטומט המקורי, כל מה שצריך זה רק לשנות את הגדרת δ שתבצע את הסימולציה כנדרש.

<u>הוכחה</u>:
יהי הומומורפיזם $h: \Sigma \to \Delta^*$, ותהי $L \subseteq \Delta^*$ שפה רגולרית.
קיים אס"ד $A = (Q, \Delta, q_0, \delta, F)$ כך ש-$L(A) = L$.
נבנה אס"ד A' באופן הבא:

$$A' = (Q, \Sigma, q_0, \delta', F)$$
$$\forall q \in Q, \sigma \in \Sigma: \quad \delta'(q,\sigma) = \hat{\delta}(q, h(\sigma))$$

נוכיח שמתקיים: $L(A') = h^{-1}(L)$, ומכאן שהשפה $h^{-1}(L)$ היא שפה רגולרית.

<u>טענת עזר</u>: $\forall q \in Q, w \in \Sigma^*: \quad \widehat{\delta'}(q,w) = \hat{\delta}(q, h(w))$
הוכחה באינדוקציה על $|w|$:
<u>בסיס</u>: $|w| = 0$ כלומר $w = \epsilon$. אז מתקיים: $\widehat{\delta'}(q,\epsilon) = q = \hat{\delta}(q,\epsilon) = \hat{\delta}(q, h(\epsilon))$, כנדרש. (שני המעברים הראשונים ע"פ הגדרת הרחבת פונקצית המעברים באס"ד, השלישי ע"פ הגדרת הרחבת h למילים).
<u>צעד</u>: $|w| = n+1$ ונניח שהטענה מתקיימת לכל מילה u כך ש-$|u| \leq n$.
$|w| \geq 1$, אז נסמן $w = u\sigma$ כך ש-$\sigma \in \Sigma, u \in \Sigma^*$. מתקיים:

$$\widehat{\delta'}(q,w)=\widehat{\delta'}(q,u\sigma)=\delta'\left(\widehat{\delta'}(q,u),\sigma\right)=\delta'\left(\hat{\delta}\big(q,h(u)\big),\sigma\right)$$
$$=\hat{\delta}\left(\hat{\delta}\big(q,h(u)\big),h(\sigma)\right)=\hat{\delta}\big(q,h(u)h(\sigma)\big)=\hat{\delta}(q,h(u\sigma))$$
$$=\hat{\delta}(q,h(w))$$
כנדרש. (המעבר השני לפי הגדרת $\widehat{\delta'}$, השלישי לפי הנחת האינדוקציה עלuu , הרביעי ע״פ הגדרת δ', החמישי ע״פ משפט מההרצאה, השישי ע״פ הרחבת h למילים).

כעת, מטענת העזר נובע שמתקיים:
$$L(A')=\left\{w\middle|\widehat{\delta'}(q_0,w)\in F\right\}=\left\{w\middle|\hat{\delta}(q_0,h(w))\in F\right\}=\{w|h(w)\in L\}=h^{-1}(L)$$
(המעברים הראשון והשלישי מהגדרת שפה של אס״ד, השני מטענת העזר, הרביעי מהגדרת h^{-1}).

5. יחסים ואפיון אלגברי של שפות.
הוכיחו/הפריכו את הטענות הבאות:
א. אם קיים יחס שקילות, R, המעדן שפה L וכן index(R) הוא סופי אז L שפה רגולרית.
ב. אם קיים יחס שקילות, R, אינווריאנטי מימין המעדן שפה L וכן $index(R)$ הוא אינסופי אז L שפה לא רגולרית.

<u>פתרון:</u>
א. לא נכון! ניקח לדוגמא את היחס R שמתקבל ע״י החלוקה: $\{L,\bar{L}\}$. המחלקות כמובן זרות ומכסות את Σ^*, לכן מגדירות יחס שקילות.
כמו-כן, אם xRy אז מהגדרת המחלקות $x,y\in L$ או $x,y\notin L$, כלומר R מעדן את L ע״פ הגדרת עידון שפה. כמו-כן, $index(R)\le 2$ (זה לא שוויון כי יתכן שאחת המחלקות שהגדרנו ריקה), כלומר סופי.
כל זה מתקיים ללא תלות ב-L עצמה, לכן כל שפה L לא רגולרית תהווה פה דוגמא נגדית, למשל $L=\{a^nb^n|n\ge 0\}$.
ב. לא נכון! ניקח לדוגמא את היחס R המוגדר ע״י החלוקה: $\big\{\{w\}|w\in\Sigma^*\big\}$ (כלומר כל מחלקה היא מילה אחת). המחלקות כמובן זרות ומכסות את Σ^*, לכן מגדירות יחס שקילות.
מתקיים $index(R)=|\Sigma^*|=\infty$.
כמו-כן, היחס הוא אינווריאנטי מימין כי אם xRy אז $x=y$ ואז לכל $z\in\Sigma^*$ מתקיים $xz=yz$ ולכן $xzRyz$ כנדרש.
יתר על כך, היחס R מעדן כל שפה L כי אם xRy אז $x=y$ ואז כמובן שמתקיים $(x,y\in L)\vee(x,y\notin L)$. לכן כל שפה L רגולרית תהווה פה דוגמא נגדית, למשל $\emptyset$.

6. הגדרת R_L ומשפט נרוד.
יהי $\Sigma=\{\sigma_1,\dots,\sigma_n\}$ א״ב, ונגדיר שפה L מעל א״ב Σ בצורה הבאה:
$$L=\{w\in\Sigma^*:\forall 1\le i\le n,\ \text{פעמים } i \text{ לכל היותר } w - \text{מופיעה ב } \sigma_i\}$$
<u>הגדירו</u> את מחלקות השקילות של היחס R_L, <u>וקבעו</u> האם L רגולרית. <u>הוכיחו</u> את תשובתכם.

פתרון:

<u>אינטואיציה</u>: כל מילה שיש בה אות σ_i שמופיעה יותר מ-i פעמים נמצאת במחלקה S_{out} (לא משנה איזה סיפא נוסיף, המילה לא בשפה). מבין שאר המילים, מה שמעניין אותנו זה כמה פעמים בדיוק מופיעה כל אות (לפי זה ניתן לדעת לכל סיפא אם היא מכניסה את המילה לשפה או לא), סדר האותיות אינו רלוונטי.

<u>מחלקות השקילות</u>:

לכל וקטור $\boldsymbol{v} \in \mathbb{N}^n$ כך ש-$\forall i\ , 0 \le v_i \le i$, נגדיר: $S_v = \{w \in \Sigma^*: \forall i\ , \#_{\sigma_i}(w) = v_i\}$.

בנוסף נגדיר: $S_{out} = \{w \in \Sigma^*: \exists i, \#_{\sigma_i}(w) > i\}$.

מחלקות השקילות של R_L הן S_{out} וכל S_v הנ"ל. מספר הדרכים לבחור v כנ"ל: $\prod_{i=1}^{n}(i+1) = (n+1)!$, לכן נקבל $index(R_L) = (n+1)! + 1$, ולכן L רגולרית ע"פ משפט נרוד.

<u>הוכחת מחלקות השקילות</u>:

- אף מחלקה אינה ריקה: $\sigma_1\sigma_1 \in S_{out}$, ולכל v כנ"ל: $\sigma_1^{v_1} \cdots \sigma_n^{v_n} \in S_v$.
- המחלקות מכסות את Σ^*: לכל $w \in \Sigma^*$ נסתכל על $\boldsymbol{v} = \left(\#_{\sigma_1}(w), \dots, \#_{\sigma_n}(w)\right)$. אם $\boldsymbol{v}$ מקיים את התנאי $\forall i\ , 0 \le v_i \le i$, אז $w \in S_{\boldsymbol{v}}$. אחרת, $\exists i, \#_{\sigma_i}(w) > i$ ולכן $w \in S_{out}$.
- המחלקות זרות: אין צורך להוכיח (נובע מההוכחה בהמשך שלכל x, y שלא באותה מחלקה מתקיים $(x,y) \notin R_L$, כי R_L רפלקסיבי).
- לכל x, y שהם באותה מחלקה, נחלק למקרים:
  - $x, y \in S_{\boldsymbol{v}}$. תהי סיפא $z \in \Sigma^*$. נשים לב שלכל אות מתקיים $\#_{\sigma_i}(x) = \#_{\sigma_i}(y)$ (מהגדרת $S_{\boldsymbol{v}}$) ולכן גם $\#_{\sigma_i}(xz) = \#_{\sigma_i}(yz)$. לפי הגדרת L, מס' הפעמים שכל אות מופיעה במילה קובע חד משמעית אם המילה בשפה או לא, לכן מתקיים $xz, yz \in L$ או $xz, yz \notin L$. לכן, מתקיים $(x,y) \in R_L$, כנדרש.
  - $x, y \in S_{out}$. אז קיימת σ_i כך ש-$\#_{\sigma_i}(x) > i$, וקיימת σ_j כך ש-$\#_{\sigma_j}(y) > j$. תהי $z \in \Sigma^*$, נשים לב שמתקיים גם $\#_{\sigma_i}(xz) > i$, וכן $\#_{\sigma_j}(yz) > j$. כלומר: $xz, yz \notin L$. לכן, מתקיים $(x,y) \in R_L$, כנדרש.
- לכל x, y שהם לא באותה מחלקה, נחלק למקרים:
  - $x \in S_v, y \in S_{v'}$ כך ש-$v \neq v'$. אז נבחר i כך ש-$v_i \neq v_i'$, ונניח בה"כ: $v_i > v_i'$.

    נביט בסיפא $z = \sigma_i^{i - v_i'}$ (החזקה חוקית כי בהכרח $v_i' \le i$, אחרת $\boldsymbol{v}'$ אינו וקטור חוקי להגדרת המחלקות). נשים לב שמתקיים: $\#_{\sigma_i}(xz) = v_i + i - v_i' > i$ ולכן $xz \notin L$.

    לעומת זאת, $\#_{\sigma_i}(yz) = v_i' + i - v_i' = i \le i$,

    וגם לכל אות אחרת σ_j: $\#_{\sigma_j}(yz) = \#_{\sigma_j}(y) = v_j' \le j$ (אחרת $\boldsymbol{v}'$ אינו וקטור חוקי להגדרת המחלקות), לכן $yz \in L$.

    לפיכך, $(x,y) \notin R_L$ כנדרש.
  - $x \in S_v, y \in S_{out}$. נבחר $z = \epsilon$. לכל אות σ_i מתקיים $\#_{\sigma_i}(x) = v_i \le i$ (אחרת $\boldsymbol{v}$ אינו וקטור חוקי להגדרת המחלקות) לכן $x\epsilon = x \in L$.

    לעומת זאת, מהגדרת S_{out}, קיימת σ_j כך ש-$\#_{\sigma_j}(y) > j$. לכן $y\epsilon = y \notin L$.

    לפיכך, $(x,y) \notin R_L$, כנדרש.

7. משפט נרוד.

יהי $\Sigma = \{\sigma_1, \dots, \sigma_n\}$ א״ב, $n \geq 2$, ויהי k טבעי נתון.
נגדיר את השפה הבאה:

$$L = \left\{ w \in \Sigma^* : \#_{\sigma_1}(w) \leq \#_{\sigma_2}(w) \leq \cdots \leq \#_{\sigma_n}(w) \leq k \right\}$$

קבעו אם השפה רגולרית, והוכיחו באמצעות משפט נרוד.

פתרון:

נוכיח ש-$index(R_L)$ הוא סופי, ולכן על פי משפט נרוד השפה L הינה רגולרית. נשים לב שלשם כך לא צריך לזהות במדויק את מחלקות השקילות של R_L, מספיק לתת חסם עליון סופי על מספרן.
נשים לב שהקבוצה $S = \{w \in \Sigma^* : \exists 1 \leq i \leq n, \#_i(w) > k\}$ מוכלת כולה במחלקת שקילות אחת של R_L: לכל שתי מילים $x, y \in S$, ולכל סיפא $z \in \Sigma^*$ מתקיים $xz, yz \notin L$. זאת מכיוון שלא משנה מה נוסיף להן עדיין תהיה אות שמופיעה יותר מ-k פעמים ולכן התנאי של השפה לא יתקיים. לכן $xR_L y$.
כמו-כן, נשים לב שב-$\bar{S}$ יש מס׳ סופי של מילים, שכן במילות $\bar{S}$ כל אות מופיעה לכל היותר k פעמים ולכן אורך המילים חסום ע״י kn. ראינו בכיתה שאם אורך המילים בשפה חסום אז השפה היא סופית. כלומר, במקרה הגרוע ביותר, מחלקות השקילות של R_L הן אחת שמכילה את S, ו-$|\bar{S}|$ נוספות שמכילות כל אחת מילה אחת (לא יתכנו יותר כי מחלקות השקילות אינן ריקות), ולכן קיבלנו $index(R_L) \leq 1 + |\bar{S}|$, ומכיוון ש-$|\bar{S}|$ סופי אז $index(R_L)$ הוא סופי.

8. למת הניפוח לשפות ח״ה.

תהי $L = \left\{ a^i b^j c^k \middle| 10 \leq i \leq j \leq k \right\}$.
הוכיחו באמצעות למת הניפוח לשפות ח״ה כי L אינה ח״ה.

פתרון:

- נניח בשלילה שהשפה L ח״ה, ולכן מקיימת את למת הניפוח לשפות ח״ה. יהי n הקבוע המובטח מהלמה.
- נסמן $m = \max(n, 10)$, ונתבונן במילה $z = a^m b^m c^m$. מתקיים $z \in L$ כי $m \geq 10$, וכן $|z| = 3m \geq n$, לכן קיים פירוק $z = uvwxy$ המקיים את תנאי הלמה.
- מכיוון ש- $|vwx| \leq n \leq m$, אז תת-המילה vwx יכולה להכיל לכל היותר אחת מהאותיות a, c. בפרט גם במילה vx מופיעה לכל היותר אחת מהאותיות a או c. נחלק לשני מקרים:
  - אם $\#_a(vx) > 0$, אז כאמור $\#_c(vx) = 0$. נסתכל על הניפוח: $z_2 = uv^2wx^2y \in L$. נשים לב כי מתקיים: $\#_c(z_2) = \#_c(uvwxy) + \#_c(vx) = \#_c(z) = m$. אך מצד שני, באותו אופן נקבל: $\#_a(z_2) = \#_a(uvwxy) + \#_a(vx) > m$, בסתירה לכך ש- $z_2 \in L$ (כי אז חייב להתקיים: $\#_a(z_2) \leq \#_c(z_2)$).
  - אם $\#_a(vx) = 0$, אז מכיוון ש-$|vx| \geq 1$ (מובטח מהלמה) מתקיים $\#_b(vx) > 0$ או $\#_c(vx) > 0$ (אלו כל האותיות בא״ב). נסתכל על הניפוח: $z_0 = uv^0wx^0y = uwy \in L$. נשים לב כי מתקיים: $\#_a(z_0) = \#_a(uvwxy) - \#_a(vx) = \#_a(z) = m$. אך מצד שני, באותו אופן נקבל: $\#_b(z_0) < m$ או לחלופין $\#_c(z_0) < m$, שניהם בסתירה לכך ש- $z_0 \in L$ (כי אז חייב להתקיים: $\#_a(z_0) \leq \#_b(z_0) \leq \#_c(z_0)$).

$\Leftarrow$ בכל מקרה הגענו לסתירה, לכן L אינה ח״ה.

9. למת הניפוח לשפות ח"ה.
הוכיחו את למת הניפוח לשפות ח"ה. לצורך ההוכחה, נתון דקדוק ח"ה $G = (V,T,P,S)$ שכל חוקיו הם מהצורה $A \to \alpha$ כך ש-$A \in V$ ו-$\alpha \in (V \cup T)^*$, $2 \le |\alpha| \le 5$.
<u>שימו לב</u>: יש להשתמש אך ורק בדקדוק הנתון G (ואסור, לדוגמה, להעביר אותו לצורה נורמלית).

<u>פתרון:</u>
תהי שפה ח"ה L ודקדוק ח"ה $G = (V,T,P,S)$ כנ"ל כך ש-$L(G) = L$.

<u>טענה</u>: יהי h גובה עץ הגזירה של z - מילת חזית של עץ גזירה ב-G המושרש במשתנה A כלשהו. מתקיים: $h \ge \log_5|z|$, או במילים אחרות, $5^h \ge 5$.

<u>הוכחה באינדוקציה על h - גובה עץ הגזירה:</u>

<u>בסיס</u>: $h = 0$. אזי בעץ יש רק שורש שהוא משתנה, וזו גם מילת החזית. מתקיים כנדרש: $|z| = 1 = 5^h$.
<u>צעד</u>: נניח נכונות עבור h. ויהי $h + 1$ גובה עץ הגזירה של z - מילת חזית של עץ גזירה ב-G המושרש במשתנה A. $h + 1 \ge 1$, ונתבונן בצעד הגזירה הראשון: $A \Longrightarrow \alpha$.
לכל משתנה ב-α מתקיים כי תת העץ המושרש במשתנה, גובהו לכל היותר h ולכן, על פי הנחת האינדוקציה, אורך מילת החזית שלו היא לכל היותר 5^h. בנוסף, $|\alpha| \le 5$, ולכן אורכה של z הוא לכל היותר $5 \cdot 5^h = 5^{h+1}$. מש"ל טענת העזר.

כעת, נראה שעבור $\boldsymbol{n = 5^{|V|+1}}$ מתקיימים תנאי הלמה.

תהי מילה z כך ש-$|z| \ge n$ ו-$z \in L = L(G)$, ויהי עץ גזירה ל-z ב-G. מהטענה נובע שגובה העץ מקיים: $h \ge \log_5|z| \ge \log_5 n = |V| + 1$. כלומר ישנו מסלול מהשורש לאחד העלים העובר בדרך בלפחות $|V| + 1$ צמתים פנימיים (כולל השורש, לא כולל העלה). נסתכל על $|V| + 1$ הצמתים הפנימיים הקרובים יותר לעלה במסלול הזה. כל צומת פנימי מייצג משתנה, ולכן מעקרון שובך היונים קיים משתנה שמופיע לפחות פעמיים, נסמנו A. נסמן ב-T_1 את תת-העץ ששורשו ה-A העליון, וב-T_2 את תת-העץ ששורשו ה-A התחתון. נשים לב שגובהו של T_1 הוא לכל היותר $|V| + 1$ (נשתמש בזה בהמשך). כעת, נסמן ב-w את מילת החזית של T_2, ב-vwx את מילת החזית של T_1 (w חייבת להיות תת-מילה שלה כי T_2 הוא תת עץ של T_1), וב-$uvwxy$ את מילת החזית של עץ הגזירה כולו (vwx חייבת להיות תת-מילה שלה, כנ"ל). כלומר מצאנו פירוק $z = uvwxy$, ונוכיח שהוא מקיים את תנאי הלמה:

1) T_1 הוא כאמור בגובה $h \le |V| + 1$. טענת העזר הנ"ל תקפה גם עליו (היא מתייחסת לכל עץ גזירה של G ולא רק לעץ ששורשו S). לכן $h \ge \log_5 |vwx|$. כלומר, מתקיים: $|vwx| \le 5^h \le 5^{|V|+1} = n$.
2) לשורש של T_1 יש לפחות 2 בנים, אחד מהם אב-קדמון של T_2, נסמן את תת העץ המושרש בו ב-T_3, ובן נוסף שנסמן את תת העץ המושרש בו ב-T_4 (הבן הנ"ל הוא משתנה או טרמינל σ). נשים לב ש-w תת-מילה של מילת החזית של T_3, לכן מילת החזית של T_4, נסמנה v', היא תת-מילה של v או של x (תלוי אם T_4 משמאל או ימין של T_3). נשים לב כי $v' \neq \epsilon$ (כי אין כללי ϵ). לכן נקבל: $|vx| \ge |v'| > 0$.
3) נשים לב שמתקיים $S \Rightarrow^* uAy \Rightarrow^* uvAxy \Rightarrow^* uvwxy$, ובפרט: $A \Rightarrow^* vAx$ וגם $A \Rightarrow^* w$. נוכיח באינדוקציה שלכל $i \ge 0$ מתקיים $S \Rightarrow^* uv^iAx^iy$, ולכן ע"י שרשור הגזירה $A \Rightarrow^* w$ נקבל $S \Rightarrow^* uv^iwx^iy$, ומכאן $uv^iwx^iy \in L = L(G)$, כנדרש.
<u>בסיס</u>: $i = 0$, ראינו קודם שמתקיים: $S \Rightarrow^* uAy = uv^0Ax^0y$.
<u>צעד</u>: אם $S \Rightarrow^* uv^iAx^iy$ אז ע"י שרשור הגזירה $A \Rightarrow^* vAx$ נקבל $S \Rightarrow^* uv^{i+1}Ax^{i+1}y$

10. ביטויים רגולריים, למת הניפוח, דקדוקים.

יהי Σ א״ב שאיננו מכיל את הסימנים $(,)$, $\emptyset, \epsilon, \cdot, +, *$, ותהי R_Σ קבוצת הביטויים הרגולריים מעל Σ.

נסתכל על R_Σ כשפה מעל א״ב $\Delta = \Sigma \cup \{\emptyset, \epsilon, \cdot, +, *, (,)\}$.

א. <u>הגדירו</u> את R_Σ.

ב. <u>הוכיחו/הפריכו :</u> R_Σ רגולרית.

ג. <u>הוכיחו/הפריכו :</u> R_Σ ח״ה.

<u>הנחיה :</u> כדי להוכיח יש לבנות דקדוק ח״ה ולנמק נכונותו, וכדי להפריך יש להשתמש בלמת הניפוח לשפות ח״ה.

<u>פתרון :</u>

<u>הערה</u> : אם רוצים לדייק בפתרון שאלה זו, יש לציין שבמהלך הפתרון נסמן את ϵ המילה הריקה בסימון חדש $\bar{\epsilon}$, זאת כדי שניתן יהיה להבחין בינה לבין האות $\epsilon \in \Delta$. כמו-כן נשתמש ב-[ ] בתור סוגריים המתארים קדימויות, כדי להבדיל מסימני הסוגריים בא״ב. עם זאת, אלה לא עניינים מהותיים לשאלה ופתרון שלא היה מתייחס לכך היה מתקבל.

א. יש לתת כאן במדויק את ההגדרה הסינטקטית מהכיתה. אין צורך בהגדרה הסמנטית.

ב. לא נכון!

- נניח בשלילה ש-R_Σ רגולרית, ולכן מקיימת את למת הניפוח. יהי n הקבוע המובטח מהלמה.
- נבנה באופן אינדוקטיבי ב״ר רגולרי r_i באופן הבא : $r_0 = \emptyset, r_{i+1} = (r_i + \emptyset)$. נשים לב שהשתמשנו רק בכללים של R_Σ, לכן $r_n \in R_\Sigma$.
- נראה שמתקיים $r_n = (^n\emptyset[+\emptyset)]^n$, באינדוקציה על n : בבסיס, $r_0 = \emptyset = (^0\emptyset[+\emptyset)]^0$ כנדרש. צעד, $r_{n+1} = (r_n + \emptyset) = ((^n\emptyset[+\emptyset)]^n + \emptyset) = (^{n+1}\emptyset[+\emptyset)]^{n+1}$, כנדרש.
- מתקיים $|r_n| = 4n + 1$, ולכן לפי הלמה קיים פירוק $r_n = uvw$ המקיים את תנאיה. נשים לב כי n האותיות הראשונות של r_n הן כולן "(", ומכיוון ש-$|uv| \leq n$ אז הפירוק הוא בהכרח מהצורה הבאה : $u = (^s,\ v = (^t,\ w = (^{n-s-t}\emptyset[+\emptyset)]^n$. ע״פ הלמה בהכרח מתקיים $uv^0w \in R_\Sigma$. אבל: $uv^0w = (^{n-t}\emptyset[+\emptyset)]^n$. ולכן מתקיים $\#_([uv^0w] = n - t < n = \#_)[uv^0w]$, זאת בסתירה לכך שלכל $r \in R_\Sigma$ מתקיים $\#_([r] = \#_)[r]$. טענה זו הוכחה בתרגול (ולא בהרצאה), ולכן אין להשתמש בה ללא הוכחה.

ג. נכון!

נסמן $\Sigma = \{\sigma_1, \ldots, \sigma_k\}$, ונבנה דקדוק חסר הקשר ל-R_Σ באופן הבא :

$$G = (\{S\}, \Delta, P, S)$$

$$P: \quad S \to \emptyset|\epsilon|\sigma_1| \ldots |\sigma_k|(S + S)|(S \cdot S)|(S^*)$$

הסבר : כללי הגזירה של S תואמים בדיוק את כל כללי הבסיס והיצירה של R_Σ ולכן ניתן לגזור ממנו בדיוק את כל הביטויים הרגולריים ב-R_Σ (גזירה של ביטוי r תתבצע בסדר הפוך ליצירה שלו ע״פ כללי R_Σ).

11. בעיות הכרעה, ביטויים רגולריים.

<u>תארו</u> אלגוריתם הפותר את בעיית ההכרעה הבאה. <u>הוכיחו</u> את תשובתכם. בהינתן ביטוי רגולרי r, האלגוריתם קובע האם קיימת $z \in L[r]$ עבורה קיים פירוק $z = uvw$ המקיים $|v| \geq 1$, ו- $uv^iw \in L[r]$ לכל $i \geq 0$.

<u>פתרון:</u>

<u>טענה</u>: שפה L רגולרית היא אינסופית אם ורק אם קיימת $z \in L$ עבורה קיים פירוק $z = uvw$ המקיים $|v| \geq 1$, ו- $uv^iw \in L$ לכל $i \geq 0$.

<u>הוכחה</u>:

$\Leftarrow$ אם L רגולרית ואינסופית אז אורך המילים בה אינו חסום (הוכח בכיתה). לכן קיימת מילה $z \in L$ כך ש- $|z| \geq n$ עבור n הקבוע המובטח בלמת הניפוח. מהלמה נובע ש-z מקיימת את הנדרש.

$\Rightarrow$ אם קיימת $z \in L$ המקיימת את הנדרש אז $uv^iw \in L$ לכל $i \geq 0$. מכיוון ש-$v \neq \epsilon$ מדובר באינסוף מילים שונות (כל אחת באורך שונה), לכן L אינסופית.

כעת מהטענה נובע שהאלגוריתם המבוקש צריך, בהינתן ב״ר r, להכריע האם $L[r]$ אינסופית.

בכיתה נלמד אלגוריתם עבור כך המשתמש בפונקציה רקורסיבית $infinity(r)$, לכן פשוט נשתמש באלגוריתם זה.

12. שקילות מודלים – אוטומט מחסנית.

בשאלה זו נציג מודל אוטומט מחסנית חדש.

יהי $M = (Q, \Sigma, \Gamma, \delta, q_0, \perp, F)$ אוטומט מחסנית המקבל ע״י ריקון המקיים את התנאי הבא:

לכל $q \in Q, \sigma \in \Sigma \cup \{\varepsilon\}, Z \in \Gamma$: אם $(q', \gamma) \in \delta(q, \sigma, Z)$ אז $|\gamma| \leq 2$.

אפיינו במדויק את השפות הניתנות לזיהוי באמצעות אוטומט במודל זה. <u>ציינו והוכיחו</u>:

1. האם מתקבלות בדיוק השפות הרגולריות?

<u>או</u>

2. האם מתקבלות בדיוק השפות חסרות-ההקשר?

<u>הערה</u>: אם אתם בונים אוטומט סופי או אוטומט מחסנית לצורך ההוכחה, עליכם לתת בניה מדויקת ופורמלית, וכן לתת הסבר לא פורמלי שלה. אין צורך לתת הוכחת נכונות פורמלית.

<u>פתרון:</u>

השפות הניתנות לזיהוי ע״י מודל זה הן בדיוק השפות חסרות ההקשר. נוכיח שקילות למודל אוטומט מחסנית המקבל ע״י ריקון שנלמד בכיתה (ראינו שמודל זה מקבל בדיוק את השפות חסרות ההקשר).

<u>מודל חדש ״$\subseteq$״ אוטומט מחסנית המקבל ע״י ריקון</u>: יהי M אוטומט במודל החדש. המודל החדש הינו אוטומט מחסנית בתוספת מגבלה על δ, לכן ניתן לקחת את M עצמו בתור אוטומט במודל אוטומט מחסנית המוכר. בשני המודלים הקבלה ע״י ריקון, ולכן השפה היא אותה שפה.

<u>אוטומט מחסנית המקבל ע״י ריקון ״$\subseteq$״ מודל חדש</u>: יהי M אוטומט מחסנית. נבנה אוטומט מחסנית M' במודל החדש כך שיתקיים $L_e(M') = L_e(M)$.

אינטואיציה לבניה:

נסמלץ את ריצת האוטומט M. כדי להתגבר על המגבלה שלא ניתן לרשום יותר מ-2 אותיות למחסנית בבת אחת, נכניס את האותיות שאנחנו רוצים למחסנית אחת-אחת. למשל אם יש מעבר שצריך לרשום למחסנית ABC, נעשה זאת ב-3 צעדים: הראשון ירשום למחסנית C, השני BC (במקום C), השלישי AB (במקום B). את כל הצעדים האלה נעשה כמסעי אפסילון ממצב מיוחד שזו אחריותו. כאשר המצב מסיים להכניס למחסנית את כל מה שרצינו, הוא עובר למצב המקורי שאליו האוטומט היה אמור לעבור, והסימולציה נמשכת. כדי לזכור בכל רגע מה נותר להכניס למחסנית ולאן צריך לעבור בסוף, ניצור מצב כזה לכל קומבינציה אפשרית של $q \in Q, \gamma \in \Gamma^*$ (במילים אחרות, "נקודד" לתוך שם המצב את מה שצריך לזכור).
אך זה יוצר בעיה: מספר הקומבינציות האלו הוא אינסופי, ומספר המצבים חייב להיות סופי. כדי לפתור את הבעיה, נשים לב שאין צורך בכל $\gamma \in \Gamma$ אפשרית, מספיק להגביל את אורכה של γ לאורך הכי גדול שמופיע בכלל של δ: $(p, \gamma) \in \delta(q, \sigma, Z)$ (האורך הזה חסום בהכרח מכיוון שהפלט של $\delta(q, \sigma, Z)$ תמיד סופי, לפי הגדרת אוטומט מחסנית, כלומר סך כללי המעבר ש-δ מגדירה הוא סופי).

בניה פורמלית:

יהי $M = (Q, \Sigma, \Gamma, \delta, q_0, \bot, \emptyset)$.
נסמן $K = \max\{|\gamma| : \exists\, q, p, \sigma, Z: (p, \gamma) \in \delta(q, \sigma, Z)\}$.
ונגדיר:

$M' = (Q \cup Q', \Sigma, \Gamma, \delta', q_0, \bot, \emptyset)$

$Q' = \{q_\gamma \mid q \in Q, \gamma \in \Gamma^*, |\gamma| \leq K\}$

(1) $\forall q \in Q, \sigma \in \Sigma \cup \{\epsilon\}, Z \in \Gamma:$

$\delta'(q, \sigma, Z) = \{(p, \epsilon) \mid (p, \epsilon) \in \delta(q, \sigma, Z)\} \cup \{(p_\gamma, Y) \mid (p, \gamma Y) \in \delta(q, \sigma, Z), Y \in \Gamma\}$

(2) $\forall q_{\gamma Y} \in Q', Y \in \Gamma, Z \in \Gamma: \quad \delta'(q_{\gamma Y}, \epsilon, Z) = \{(q_\gamma, YZ)\}$

(3) $\forall q_\epsilon \in Q', Z \in \Gamma: \quad \delta'(q_\epsilon, \epsilon, Z) = \{(q, Z)\}$

הסבר לבניה:

המצבים ב-Q מבצעים סימולציה ל-M, המצבים ב-Q' אחראים על הכתיבה למחסנית. המצב q_γ משמעו שיש לרשום כעת γ למחסנית ולבסוף לעבור למצב q להמשך הסימולציה.
כלל (1) של δ': כל כלל מעבר שרושם ϵ למחסנית נשאיר כמו שהוא. לכל כלל מעבר $(p, \gamma Y) \in \delta(q, \sigma, Z)$, (כלומר כלל הרושם לפחות אות אחת למחסנית), במקום לרשום γY למחסנית (לא אפשרי אם $|\gamma Y| > 2$) נרשום רק Y ונעבור למצב q_γ שאחראי לרשום את שאר γ.
כלל (2) של δ': אם אנחנו במצב $q_{\gamma Y}$ אז נוסיף Y למחסנית (על גבי Z שנמצא שם כרגע) ונעבור למצב q_γ (כלומר "נסיר" את Y מהזכרון של המצב).
כלל (3) של δ': אם אנחנו במצב q_ϵ משמע שסיימנו לרשום למחסנית מה שהיה צריך וניתן לקפוץ למצב q להמשך הסימולציה (המחסנית נותרת ללא שינוי).

<u>שימו לב</u>: מפתה לרשום את כלל (1) כך: $\delta'(q, \sigma, Z) = \{(p_\gamma, \epsilon) \mid (p, \gamma) \in \delta(q, \sigma, Z)\}$, כלומר לאפשר למצב p_γ לבצע את כל הרישום למחסנית ובכך לטפל בכל המקרים בצורה אחידה ואלגנטית. אך נשים לב שזה לא יעבוד כצפוי: המעבר ל-p_γ רושם ϵ למחסנית והוא עלול לרוקן אותה בניגוד למה שהמעבר שהוא מסמלץ היה עושה. במצב זה החישוב "יתקע" ולא ימשיך כצפוי. לכן בכל מצב ש-$\gamma \neq \epsilon$ אנחנו חייבים לרשום מיד לפחות אות אחת למחסנית כדי להשאיר אותה לא ריקה.

13. צורות נורמליות ודקדוקים ח״ה.

בהינתן שפה ח״ה L מעל א״ב Σ, נגדיר את השפה הבאה: $L' = \{\, u\,\sigma\,v \mid \sigma \in \Sigma\,,\; u,v \in \Sigma^*\,,\; uv \in L\,\}$.

א. הוכיחו ש- L' ח״ה באמצעות תכונות סגור בלבד.

ב. בהינתן דקדוק ח״ה עבור L, תנו בנייה פורמלית של דקדוק ח״ה עבור L'. אין להסתמך על סעיף א׳, אלא יש להגדיר דקדוק בצורה ישירה ומפורשת (הערה: אסור להשתמש באוטומטי מחסנית). נמקו את נכונות הבנייה.

<u>פתרון:</u>

א. תהי L שפה ח״ה.

- נגדיר הומומורפיזם $h: \Sigma \cup \{\$\} \to \Sigma^*$ באופן הבא: $h(\$) = \epsilon$, $h(\sigma) = \sigma\; \forall\sigma \in \Sigma$. השפה $h^{-1}(L)$ ח״ה מסגירות השפות ח״ה להומומורפיזם הפוך. נשים לב שהשפה $h^{-1}(L)$ היא שפת כל המילים שמתקבלות ע״י מילה מ-L ש״דחפו״ לתוכה סימני \$ בכל מיני מקומות (יתכן גם כמה ברצף), כלומר:
$$h^{-1}(L) = \left\{\$^{i_0}\sigma_1\$^{i_1}\sigma_2\$^{i_2} \ldots \sigma_n\$^{n} : n \geq 0, \sigma_1\sigma_2\cdots\sigma_n \in L\right\}$$
- השפה $\Sigma^*\{\$\}\Sigma^*$ היא רגולרית (מסגירות השפות הרגולריות לשרשור וכי $\{\$\}$ סופית ולכן רגולרית, ו-Σ^* רגולרית). לכן השפה $L_2 = h^{-1}(L) \cap \Sigma^*\{\$\}\Sigma^*$ הינה ח״ה מסגירות השפות ח״ה לחיתוך עם שפה רגולרית. נשים לב שמתקיים: $L_2 = \{u\$v \mid u,v \in \Sigma^*, h(u\$v) \in L\} = \{u\$v \mid uv \in L\}$.
- נגדיר הצבה $f: \Sigma \cup \$ \to 2^{\Sigma^*}$ באופן הבא: $f(\$) = \Sigma$, $f(\sigma) = \{\sigma\}\; \forall\sigma \in \Sigma$. נשים לב ש-$f$ מחזירה שפה סופית בכל המקרים, לכן רגולרית ובפרט ח״ה, ולכן ההצבה f היא הצבה ח״ה. לכן $f(L_2)$ היא שפה ח״ה מסגירות השפות ח״ה להצבה ח״ה.
- מתקיים: $f(L_2) = \{f(u\$v) \mid uv \in L\} = \{u\sigma v \mid \sigma \in \Sigma, uv \in L\} = L'$. כלומר קיבלנו שהשפה L' היא ח״ה, כנדרש.

ב. <u>רעיון הבניה</u>: נשנה את כללי הדקדוק G של L כך שיכפו גזירת אות נוספת במקום כלשהו במילה, ורק אחת. נשאיר את המשתנים וכללי הגזירה של G, ובנוסף, על כל משתנה X נגדיר משתנה X' שייעודו הוא ״לגזור מה ש-X גוזר, בתוספת אות אחת איפשהו״. בכל גזירה מ-X' נבחר אחת משתי אפשרויות: לגזור תבנית פסוקית ע״פ כלל של G ונוסיף לתבנית אות אחת, או שנעביר את האחריות להוספת האות לאחד מהמשתנים שגזרנו (ע״י כך שנתייג אותו).

<u>בניה פורמלית</u>:

יהי $G = (V, \Sigma, P, S)$ דקדוק ח״ה עבור L. נבנה את הדקדוק הבא:

$$G' = (V' \cup V, \Sigma, P \cup P', S')$$
$$V' = \{X' \mid X \in V\}$$
$$P' = \{X' \to \alpha\sigma\beta \mid \sigma \in \Sigma, X \to \alpha\beta \in P\} \cup \{X' \to \alpha Y'\beta \mid Y \in V, X \to \alpha Y\beta \in P\}$$

הערה: ניתן להשתמש כאן בצורה נורמלית של חומסקי כדי לפשט מעט את הבניה. אך יש לשים לב שאנחנו גם מטפלים במקרה ש-$\epsilon \in L$.

<u>נימוק נכונות</u>: כפי שנכתב ברעיון הבניה. ב-G' יש את אותם הכללים כמו ב-G, ובנוסף גם כללי גזירה של משתנים מתויגים בהם מכניסים בדיוק אות אחת נוספת בתבנית הנגזרת ע״י כללי P או מעבירים אחריות הלאה להמשך הגזירה – ע״י תיוג משתנה אחד בתבנית הנגזרת על פי כללי P. המשתנה ההתחלתי – מתויג על מנת להכניס אות נוספת במילה.

14. בניית אוטומט מחסנית.

בהינתן אוטומט סופי דטרמיניסטי $A = (Q, \Sigma, \delta, q_0, F)$.

א. תנו בניה של אוטומט מחסנית **בעל מצב יחיד**, $M = (\{q_M\}, \Sigma, \Gamma, q_M, \perp, \delta_M, \emptyset)$,
כך ש-$L_e(M) = L(A)$.

ב. השלימו את טענת העזר הבאה והשתמש בה על מנת להוכיח את נכונות הבניה בסעיף א'.
אין צורך להוכיח את טענת העזר.
טענת עזר: אם: $(q_M, x, \perp) \vdash^* (q_M, \varepsilon, \alpha)$, כאשר: $x \in \Sigma^*, \alpha \in \Gamma^+$, אז: $\alpha = \_\_\_\_$

<u>פתרון:</u>

א. <u>רעיון הבניה</u>: אנחנו רוצים לסמלץ אס"ד באוטומט מחסנית, לכן אין צורך להשתמש במחסנית בכלל. מצד שני, הוגבלנו לשימוש במצב אחד, לכן ננצל את המחסנית כדי לזכור באיזה מצב של האס"ד אנו נמצאים (א"ב המחסנית יהיה מצבי האס"ד, ובכל רגע נתון יהיה במחסנית תו אחד שאומר מה המצב הנוכחי).

<u>בניה פורמלית</u>:

$$M = (\{q_M\}, \Sigma, Q, q_M, q_0, \delta_M, \emptyset)$$

$$(1) \quad \forall q \in Q, \sigma \in \Sigma: \quad \delta_M(q_M, \sigma, q) = \{(q_M, \delta(q, \sigma))\}$$

$$(2) \quad \forall q_f \in F: \quad \delta_M(q_M, \epsilon, q_f) = \{(q_M, \epsilon)\}$$

נשים לב כי התו ההתחלתי במחסנית הוא q_0, ו-$\Gamma = Q$.

ב. <u>הוכחת נכונות</u>:
טענת העזר השלמה: אם $(q_M, x, q_0) \vdash^* (q_M, \epsilon, \alpha)$ כאשר $\alpha \in Q^+$, $x \in \Sigma^*$, אז $\alpha = \hat{\delta}(q_0, x)$.
(נשים לב שהחלפנו את $\perp$ ב-q_0 ואת Γ ב-Q, ע"פ הגדרת M).

כעת נוכיח $L_e(M) = L(A)$:

- תהי $x \in L_e(M)$, לכן מתקיים $(q_M, x, q_0) \vdash^* (q_M, \epsilon, \epsilon)$ ע"פ הגדרת שפה של אוטומט מחסנית המקבל ע"י ריקון. נשים לב שצעד החישוב האחרון רוקן את המחסנית, לכן הוא בהכרח ע"פ כלל (2) של δ_M (הכלל הראשון רושם תו למחסנית). כלומר מתקיים:
$(q_M, x, q_0) \vdash^* (q_M, \epsilon, q_f) \vdash (q_M, \epsilon, \epsilon)$ כאשר $q_f \in F$. מטענת העזר נקבל $\hat{\delta}(q_0, x) = q_f$ ולכן $x \in L(A)$ כנדרש.
- תהי $x \in L(A)$, אז: $\hat{\delta}(q_0, x) = q_f \in F$. נשים לב שכלל (1) של M יכול להיות מופעל על כל קונפיגורציה של M (כל עוד נותר קלט והמחסנית לא ריקה), ושהפעלתו מעולם לא מרוקנת את המחסנית (כי רושמת אליה תו), ומקצרת בכל הפעלה את הקלט בתו אחד. לכן כש-M נמצא בקונפיגורציה (q_M, x, q_0) ניתן להפעיל את הכלל $|x|$ פעמים עד שהקלט נגמר. כלומר נקבל $(q_M, x, q_0) \vdash^{|x|} (q_M, \epsilon, \alpha)$. מטענת העזר נובע ש- $\alpha = \hat{\delta}(q_0, x) = q_f \in F$, ולכן ע"י הפעלת כלל (2) נקבל $(q_M, \epsilon, q_f) \vdash (q_M, \epsilon, \epsilon)$. בסה"כ קיבלנו $(q_M, x, q_0) \vdash^* (q_M, \epsilon, \epsilon)$, כנדרש.