אוטומטים - דף נוסחאות

פעולות על שפות

$L_1 \Delta L_2 = \{w \in L_1 \ \wedge \ w \notin L_2\} \ \cup \ \{w \notin L_1 \ \wedge \ w \in L_2\}$

$L_1 \cdot L_2 = \{uv \mid u \in L_1,\ v \in L_2\}$	$L_1 \setminus L_2 = \{w \in L_1 \ \wedge \ w \notin L_2\}$
$\lvert L_1 \cdot L_2 \rvert \le \lvert L_1 \rvert \cdot \lvert L_2 \rvert$	$L \cdot \emptyset = \emptyset = \emptyset \cdot L$

דוגמה: $L_1 = \{aba,\ aa\}$, $L_2 = \{abc,\ aa\}$ אז:

$L_1 \setminus L_2 = \{aba\} \quad , \quad L_2 \setminus L_1 = \{abc\} \quad , \quad L_1 \Delta L_2 = \{aba,\ abc\}$

$L_1 \cdot L_2 = \{abaabc,\ abaaa,\ aaabc,\ aaaa\}$

$L_2 \cdot L_1 = \{abcaba,\ abcaa,\ aaaba,\ aaaa\}$

$L^* = \cup_{i=0}^{\infty} L^i = L^0 \cup L^1 \cup L^2 \cup .....$

אם $L = \Sigma$ אז $L^* = \Sigma^*$	$\lvert\Sigma^*\rvert = \aleph_0$ אינסופית
$\overline{L_1} = \Sigma^* \setminus L_1$	$L_1 \setminus L_2 = L_1 \cap \bar{L_2}$
$\varepsilon \in L_2 \leftrightarrow L_1 \subseteq L_1 \cdot L_2$	$(L_1 \cup L_2)L_3 = L_1L_3 \cup L_2L_3$

אוטומט סופי דטרמיניסטי (אס"ד - DFA)

$\delta_A: Q \times \Sigma \to Q$	$A = (\Sigma,\ Q,\ q_0,\ F,\ \delta)$

$\hat{\delta}(q, w_1w_2) = \hat{\delta}(\hat{\delta}(q, w_1),\ w_2)$

$L(A) = \{w \in \Sigma^* \mid \hat{\delta}(q_0, w) \in F\}$	$L(A) = \cup_{q \in F} L_A(q)$

שפה L רגולרית אם היא מתקבלת ע"י DFA

אוטומט DFA עם מצב מקבל יחיד לא שקול לאוטומט DFA רגיל.

שפות רגולריות

שפות סופיות תמיד רגולריות ושפות שהמשלימה שלה סופית תמיד רגולריות

סגירות לשפות רגולריות

אם L_1, L_2 רגולריות שפות רגולריות אז:

- ***איחוד***: $L_1 \cup L_2$ רגולרית. להוכיח בדה מורגן: $L_1 \cup L_2 = \overline{\overline{L_1} \cap \overline{L_2}}$.
  השפות הרגולריות אינן סגורות לאיחוד אינסופי.
- ***חיסור***: $L_1 \setminus L_2$ רגולרית. להוכיח בדיאגרמת וון: $L_1 \setminus L_2 = L_1 \cap \overline{L_2}$.
- ***שרשור***: $L_1 \cdot L_2$ רגולרית. (ניתן להראות בעזרת מסעי אפסילון).
- ***חיתוך***: $L_1 \cap L_2$ רגולרית. ניתן להוכיח ע"י אוטומט מכפלה.
- **סימטרי**: $L_1 \Delta L_2$ רגולרית. (כי משלים, חיתוך ואיחוד משמרות רגולריות).

תיאור פורמלי של אוטומט מכפלה: נבנה אוטומט מכפלה A שיחקה את $A_1,\ A_2$.
עבור אוטומט מכפלה $A = (\Sigma,\ Q,\ q_0,\ F,\ \delta)$ כאשר: $Q = Q_1 \times Q_2$. מצב התחלתי: $q_0 = (q_{01},\ q_{02})$ קבוצת מצבים מקבלים: $F = F_1 \times F_2$. מעברים:
$\forall q_1 \in Q_1,\ q_2 \in Q_2: \delta((q_1, q_2), \sigma) = (\delta_1(q_1, \sigma),\ \delta_2(q_2, \sigma))$

- **הכלה** (**לא בהכרח** רגולרית).

אם L שפה רגולרית אז גם השפות הבאות **רגולריות**: $n \in \mathbb{N}: L^n,\ \bar{L},\ L^R,\ L^*$

סגירות לשפות L ***<u>לא</u>*** רגולריות:

משלים: $\bar{L}$ לא רגולרית.	***היפוך***: L^R לא רגולרית.

אוטומט סופי אי-דטרמיניסטי (אסל"ד - NFA)

מודל NFA שקול למודל DFA ולכן כל שפה שמתקבלת בו היא גם רגולרית.

$\delta: Q \times \Sigma \to S \subseteq Q$	$A = (\Sigma,\ Q,\ q_0,\ F,\ \delta)$
$\hat{\delta}(q, wa) = \cup_{p \in \delta(q,w)} \delta(p, a)$	*בסיס:* $\delta(q, \varepsilon) = \{q\}$
$\hat{\delta}(P, w) = \cup_{q \in P} \hat{\delta}(q, w)$	

שפת האוטומט: $L(A) = \{w \in \Sigma^* \mid \hat{\delta}(q_0, w) \ \cap \ F \neq \emptyset\}$

אוטומט חזקה: נשתמש כדי להפוך NFA ל- DFA והוא יקבל את אותה השפה.
נבנה את האס"ד D הבא באמצעות האסל"ד הנתון N:

- קבוצת המצבים היא $P(Q)$. (כל תתי הקבוצות של המצבים, פרט ל-$\emptyset$).
- מצב התחלתי $\{q_0\}$.
- קבוצת מצבים מקבלים F_D היא כל תתי הקבוצות של Q המכילות איבר מ-F.
- מעברים δ_D היא $\delta_D(\{q_1, ..., q_k\}, \sigma)$ קבוצה המתקבלת מאיחוד כל $1 \le i \le k$ של המעברים מהאסל"ד $\delta_N(q_i, \sigma)$.

אוטומט סופי אי-דטרמיניסטי עם מסעי אפסילון (ENFA)

מודל ENFA שקול למודל NFA ולכן כל שפה שמתקבלת בו היא גם רגולרית.

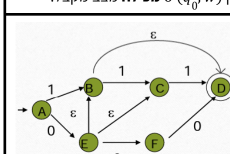
$\delta: Q \times (\Sigma \cup \{\varepsilon\}) \to P(Q)$	$A = (\Sigma,\ Q,\ q_0,\ F,\ \delta)$

סגור אפסילון של מצב q הוא: קבוצת המצבים שניתן להגיע אליהם מ-q על ידי שימוש במסעי ε בלבד ונסמן ב- $CL^{\varepsilon}(q)$.
סגור של קבוצת מצבים P הוא: $CL^{\varepsilon}(P) = \cup_{q \in P} CL^{\varepsilon}(q)$.

הרחבה למילים של פונקציית המעברים: בסיס: $\delta'(q, \varepsilon) = CL^{\varepsilon}(q)$.
צעד: $\delta'(q, x\sigma) = \cup_{p \in \delta'(q,x)} CL^{\varepsilon}(\delta(p, \sigma))$.

שפה של ENFA היא קבוצת כל המחרוזות w עבורן $\hat{\delta}'(q_0, w)$ **מכילה** מצב מקבל.

דוגמה פרקטית



$\delta'(A, \varepsilon) = CL^{\varepsilon}(A) = \{A\}$

$CL^{\varepsilon}(B) = \{B, D\}$

$CL^{\varepsilon}(C) = \{C\}$

$CL^{\varepsilon}(E) = \{E, B, C, D\}$

$\delta'(A, 0) = \{E, B, C, D\}$

$\delta'(A, 01) = \cup_{p \in \delta'(A,0)} CL^{\varepsilon}(\delta(p, 1)) = \{CD\}$

ביטויים רגולריים

אוסף הביטויים הרגולריים מעל א"ב Σ המסומן ב- R_Σ מוגדר באינדוקציה מבנית:

- **אטומים**: $\phi, \varepsilon \in R \quad , \quad \forall \sigma \in \Sigma: \sigma \in R$
- **פעולות יצירה:**
  - אם $r_1, r_2 \in R$ אז $(r_1 + r_2) \in R$ וכן $(r_1 \cdot r_2) \in R$.
  - אם $r \in R$ אז $(r^*) \in R$.

תהי $L[r]$ **השפה** שמציין את הביטוי הרגולרי r. נגדיר את הפונקציה L מ-R ל- 2^{Σ^*} באופן הבא:
$L[\phi] = \phi \quad , \quad L[\varepsilon] = \varepsilon \quad , \quad \forall \sigma \in \Sigma: L[\sigma] = \sigma$

- אם $r_1, r_2 \in R$ אז:
  - $L[(r_1 + r_2)] = L[r_1] \cup L[r_2]$.
  - $L[(r_1 \cdot r_2)] = L[r_1] \cdot L[r_2]$.
- אם $r \in R$ אז: $L[(r^*)] = (L[r])^*$.

סדר קדימויות להשמטת סוגריים: * קדימות גבוהה ביותר (איטרציה) · קדימות בינונית (שרשור) + קדימות נמוכה (איחוד)	$r^+ = (r \cdot (r^*))$ $\Sigma = \{a, b\} \to \Sigma^* = (a + b)^*$ $(\Sigma\Sigma)^* = ((a + b)(a + b))^*$

משפט: לכל $r \in R$ מעל Σ מתקיים $L[r]$ היא שפה רגולרית. (הוכחה באינדוקציה מבנית)
משפט: לכל שפה רגולרית $L \subseteq \Sigma^*$ קיים ביטוי רגולרי r כך ש: $L[r] = L$.

שקילות ביטויים רגולרים: נוכיח בהכלה דו כיוונית $L[r_1] = L[r_2]$.

אי-שקילות ביטויים רגולרים: נראה דוגמה נגדית: $w \in L[r_1]$ אבל $w \notin L[r_2]$.

שיטות להפקת ביטוי רגולרי:

1. שיטת מאוטומט לביטוי רגולרי (נבנה אוטומט ל-L, ניצור ב"ר מכל מסלול ואז נחבר):
   a. בונים לשפה אוטומט כמה שיותר אי-דטרמיניסטי (כמה שיותר מינימלי).
   b. מייצרים מצב מקבל יחיד על ידי מסעי אפסילון.
   c. בכל שלב מוחקים מצב שאינו ההתחלתי ואינו המקבל ובודקים מה המסלול באורך 2 שמגיע מכל מצב לכל מצב (ששונה ממה שמחקנו) דרך אותו מצב שמחקנו. שאר המצבים אותו הדבר.
2. שיטת הבלוקים (נפריד את חלקי המילה לבלוקים ונבדוק איזה ביטוי מותר בכל בלוק).

למת הניפוח לשפות רגולריות

- **למה**: בכל אס"ד קיים מעגל (הוכח בהרצאה - לפי שובך היונים).
- למת הניפוח הינה תנאי הכרחי ואינו מספיק לרגולריות (יש לא רגולריות שמקיימות).
- עיקר השימוש בלמה יהיה להפריך רגולריות של שפה.

תבנית להפרכת רגולריות עם למת הניפוח: נניח בשלילה כי L רגולרית, לכן היא מקיימת את למת הניפוח, לכן לפי הלמה קיים קבוע $n \in \mathbb{N}$ כך שלכל $\lvert z\rvert \ge n$ קיים פירוק מהצורה uvw כך ש: $\lvert uv\rvert \le n$ וגם $1 \le \lvert v\rvert \le n$ שמקיים עבור כל $i \in \mathbb{N}$ מילה מהשפה $uv^iw \in L$. יהא $z = \_\_ \in L$, $\lvert z\rvert = \_\_ \ge n$. נראה שקיים $i = \_\_ \in \mathbb{N}$ עבורו המילה uv^iw אינה בשפה L ובכך מגיעים לסתירה.

שפה **לא** רגולרית הניתנת לניפוח בלמה: $L = \{a\}^* \cup \{b^ja^{k^2} \mid 1 \le j, k\}$

שילוב למת הניפוח עם תכונות סגור

1. נניח בשלילה ש: L רגולרית.
2. נגדיר את L' שנוצרה מסגירות רגולרית של L על איחוד/חיתוך/משלים וכדומה..
3. נראה שהשפה L' לא מקיימת את למת הניפוח ולכן מסגירות גם L לא רגולרית.

יחסים

יחס שקילות הוא יחס R המקיים את שלושת התכונות הבאות (נשתמש להוכיח יחס שקילות):

- היחס רפלקסיבי אם $\forall x:\ R(x, x)$
- היחס סימטרי אם $\forall x, y:\ R(x, y) \to R(y, x)$
- היחס טרנזיטיבי אם $\forall x, y, z:\ R(x, y) \ \wedge \ R(y, z) \to R(x, z)$

מחלקות שקילות: יחס שקילות משרה חלוקה של העולם לתת קבוצות זרות ומשלימות.
יהי R יחס שקילות מעל הקבוצה A:

- נסמן ב-$index(R)$ את כמות מחלקות השקילות.
- נציג מתוך המחלקות יסומן $[x] = \{y:\ R(x, y)\}$.
- יחס R' **מעדן** את R אם $R' \subseteq R$ כלומר: $\forall x, y \in A:\ R'(x, y) \to R(x, y)$.
  כלומר שכל מחלקת שקילות של R' מוכלת במחלקת שקילות של R.

עידון של שפה: יהי R יחס שקילות מעל Σ^* ו-$L \subseteq \Sigma^*$ שפה כלשהי.
נאמר כי R מעדן את L אם ורק אם מתקיים: $R(x, y) \leftrightarrow (x \in L \leftrightarrow y \in L)$.
כלומר שתי מילים יהיו ביחס אם שתיהן בשפה או שתיהן לא בשפה.

יחס שקילות R מעל Σ^* ייקרא **אינווריאנטי מימין** אם ורק אם הוא מקיים:
$\forall x, y \in \Sigma^* :\ R(x, y) \to \forall z \in \Sigma^* :\ R(xz,\ yz)$

יחס שקילות R_A

יחס שקילות R_A **לאוטומט** A מוגדר כך: $R_A(x, y) \Leftrightarrow \hat{\delta}(q_0, x) = \hat{\delta}(q_0, y)$.

- באס"ד בו כל המצבים ישיגים, ל-R_A יש $\lvert Q\rvert$ מחלקות שקילות.
- R_A אינווריאנטי מימין (הוכח בהרצאה ב-3 שורות).
- **טענה**: לכל אס"ד A, היחס R_A מעדן את $L(A)$. (הוכח בהרצאה).
- בנוסף, $index(R_A) \le \lvert Q_A\rvert$ ובפרט $index(R_A) < \infty$.
- $L(A) = \cup_{qi \in F} S_i = \{x \mid \exists q \in F: \hat{\delta}(q_0, x) = q\}$

יחס שקילות R_L

יחס R_L **לשפה** L מוגדר כך: $R_L(x, y) \Leftrightarrow (\forall z \in \Sigma^*:\ xz \in L \leftrightarrow yz \in L)$.
היחס R_L מקבל זוג אם"ם לכל סיפא z, המילים xz, yz מתקבלות יחד או נדחות יחד.

- יחס R_L הוא יחס שקילות.

סדר פעולות להוכחת מחלקות השקילות של R_L יש להראות את הדברים הבאים:

1. המחלקות אינן ריקות, והן מכסות את Σ^*.
2. כל זוג מילים x, y מאותה מחלקה S אינן ניתנות להפרדה על ידי סיפא z.
3. כל זוג מילים ממחלקות שונות, כן ניתנות להפרדה על ידי סיפא z כלשהי.

משפט האפיון של R_L (הוכח בהרצאה): לכל שפה $L \subseteq \Sigma^*$ מתקיים:

1. R_L הוא יחס שקילות אינווריאנטי מימין.
2. R_L מעדן את L.
3. אם R יחס שקילות אחר, אינווריאנטי מימין המעדן את L אזי R מעדן גם את R_L.

כלומר, R_L הוא יחס שקילות אינווריאנטי מימין המכיל הכי מעט מחלקות שקילות.

משפט נרוד

שפה L רגולרית אם ורק אם $index(R_L)$ סופי

תהי $L \subseteq \Sigma^*$. אזי הטענות הבאות שקולות:

1. L רגולרית.
2. קיים יחס שקילות R אינווריאנטי מימין המעדן את L ומקיים: $index(R) < \infty$.
3. $index(R_L) < \infty$.

הוכחת אי רגולריות בעזרת משפט נרוד

1. נגדיר קבוצה אינסופית של מילים (לאו דווקא מהשפה הנתונה).
2. נרצה להראות שכל זוג מילים $x \neq y$ מהקבוצה שהגדרנו לא מקיים את היחס.
3. נבחר סיפא מפרידה z עבור זוג המילים ונראה (בה"כ) ש: $xz \in L$ אבל $yz \notin L$.
4. לכן, מספר מחלקות השקילות היא לפחות ממספר המילים בקבוצה - אינסוף מילים.
5. כלומר הראינו ש $index(R_L)$ אינסופי ולכן לפי משפט נרוד השפה L לא רגולרית.

מינימליזציה של אס"ד (Hopcroft)

כמות מחלקות השקילות של R_L יוצרות את האס"ד המינימלי ל-L

טענה (הוכח בהרצאה): אם A אס"ד המקבל את השפה L אזי $index(R_L) \le \lvert Q_A\rvert$ שהוא מספר מצבי האוטומט A_L מהוכחת משפט נרוד. $(index(R_L) \le index(R_A))$.

האלגוריתם של Hopcroft:

1. מחק מצבים שאינם ישיגים מ- q'_0. לאחר המחיקה, סמן את האוטומט המתקבל ב-A.
2. אתחל יחס $R = \emptyset$. זהו יחס סימטרי (בינארי על Q).
   a. עדכן: $R \leftarrow \{\{p, q\} \mid q \in F \wedge p \notin F\}$.
      i. חשב: $S \leftarrow \{\{p, q\} \notin R \mid \exists \sigma \in \Sigma:\ \{\delta(q, \sigma),\ \delta(p, \sigma)\} \in R\}$.
      ii. אם $S \neq \emptyset$ עדכן $R \leftarrow R \cup S$ וחזור לשלב (i).
3. אם $\bar{R}$ אינו יחס שקילות על Q החזר $Fail$.
4. בנה DFA מצומצם $\tilde{A} = (\tilde{Q},\ \Sigma,\ \tilde{q}_0,\ \tilde{\delta},\ \tilde{F})$ באופן הבא:
   a. $\tilde{Q}$= קבוצת מחלקות השקילות של $\bar{R}$.
   b. $\tilde{q}_0$= מחלקת השקילות של q_0 ב-$\bar{R}$. נסמן ב- $[q_0]_{\bar{R}}$.
   c. $\tilde{F} = \{\tilde{q} \mid \tilde{q} \cap F \neq \emptyset\}$.
   d. $\forall \tilde{q} \in \tilde{Q},\ \forall \sigma \in \Sigma:\ \tilde{\delta}(\tilde{q},\ \sigma) = [\delta(q, \sigma)]_{\bar{R}}$.
   e. כדי לדעת מה מכיל את $\bar{R}$, נבנה טבלת "הצלבות" ונסמן את כל הזוגות שנמצאות ב-R. כל משבצת שנותרה ריקה היא זוג ב-$\bar{R}$. במידה ויש מצב q בטבלה (עמודה או שורה) שכולה מסומנת חוץ מהמצב עצמו אז q הוא מחלקת שקילות בפני עצמו.

- **טענה 0**: האלגוריתם תמיד עוצר.
- **מסקנה 1**: האלגוריתם לא מחזיר Fail כלומר $\bar{R}$ שמתקבל בסוף הוא אכן יחס שקילות.
- **מסקנה 2**: $index(R_L(A)) = index(\bar{R})$. כלומר $\bar{R}$ הפריד בין שפות מצבים בדיוק כמו ש $R_L(A)$ אמור להפריד.
- **טענה 2**: אם $\tilde{q} \in \tilde{F}$ אזי $\tilde{q} \subseteq F$.

דקדוקים

דקדוק הוא **רביעייה** $G = (T, V, P, S)$ כאשר:

- V קבוצה לא ריקה וסופית של משתנים.
- T קבוצה לא ריקה וסופית של טרמינילים (אותיות) הזרה ל-V.
- P קבוצה לא ריקה וסופית של כללי שכתוב (או כללי גזירה) כך שכל כלל ב-P הוא מהצורה $\alpha \to \beta$ כאשר $\alpha \in (V \cup T)^* \cdot V \cdot (V \cup T)^*$ ו-$\beta \in (V \cup T)^*$.

בהינתן דקדוק G, **השפה** $L(G)$ מוגדרת כך: $L(G) = \{x \in T^* \mid S \Rightarrow^* x\}$.
הוכחת $L = L(G)$ היא באינדוקציה בהכלה דו-כיוונית.

- כיוון ראשון $L \subseteq L(G)$ נוכיח באינדוקציה חזקה על $\lvert w\rvert$.
- כיוון שני $L(G) \subseteq L$ נוכיח באינדוקציה על אורך הגזירה ל-γ (טענה חזקה).

ההירכיה של חומסקי

הרעיון הכללי הוא של ידי השמת "מגבלות" על דקדוקים, נוכל לסווג אותם לכמה סוגים שכל אחד מהם מתאים למודל "אוטומטים" שונה. קיימים ארבעה סוגי דקדוקים, כל קבוצה עוקבת מוסיפים עוד מגבלות: $\Gamma_3 \subseteq \Gamma_2 \subseteq \Gamma_1 \subseteq \Gamma_0$:

- Γ_1 דקדוקים תלויי הקשר: אם $\alpha \to \beta$ אזי $\lvert\alpha\rvert \le \lvert\beta\rvert$ ומותר $S \to \varepsilon$.
- Γ_2 דקדוקים חסרי הקשר: באגף שמאל תמיד יהיה משתנה בודד. כל כללי הגזירה מהצורה $\alpha \to \beta$ דורשים כי $\alpha \in V^1$. ומותר $S \to \varepsilon$. מודל זה שקול לאוטומט מחסנית אי-דטרמיניסטית.
- Γ_3 דקדוקים רגולריים: כל כללי השכתוב הם מהצורה $A \to aB$ או $A \to a$ או $S \to \varepsilon$ $(\forall A, B \in V,\ \forall \sigma \in \Sigma)$. מודל זה שקול לשפות רגולריות (DFA,NFA,...).
  נחלק את Γ_3 בין דקדוק ליניארי ימני $A \to aB$ ובין דקדוק ליניארי שמאלי $A \to Ba$.

משפט: קבוצת השפות הרגולריות מוכלת ב-Γ_3.
טענת עזר 1: יהי G דקדוק ליניארי ימני. אזי אם מתקיים $A \Rightarrow^* \alpha$ (כאשר $A \in V^1,\ \alpha \in (V \cup T)^*$) אז α מהצורה wB או w כאשר $w \in T^*$ ו-$B \in V^1$.
טענת עזר 2: $\forall w \in \Sigma^*,\ \forall p, q \in Q \ :\ \hat{\delta}(p, w) = q \leftrightarrow p \Rightarrow^* wq$.
טענה: דקדוק ליניארי ימני שקול בכוחו לדקדוק ליניארי שמאלי.

איך עוברים מדקדוק ליניארי ימני לאוטומט?

- יוצרים אוטומט אי דטרמיניסטי.
- הא"ב של האוטומט יהיה הא"ב של הדקדוק.
- המצבים יהיו המשתנים (המצב ההתחלתי יהיה המשתנה ההתחלתי).
- אם $A \to aB$ אז: $\delta(A, a) = B$.
- אם $A \to a$ אז $\delta(A, a) = q_f$.
- אם $S \to \varepsilon$ אז: $\delta(S, \varepsilon) = q_f$.
- המצב המקבל יהיה רק q_f.

סדר פעולות למעבר מדקדוק ליניארי ימני לדקדוק ליניארי שמאלי

1. לפי האוטומט של L, בנה אס"ד ל- L^R ואז בנה לפי האס"ד דקדוק ליניארי ימני ל-L^R:
2. הפוך את הכללים של הדקדוק (מצבי האוטומט): נחליף כל $q_i \to \sigma q_j$ ב-$q_i \to q_j\sigma$.
3. אם ממצב q_i עם האות σ נשלח למצב מקבל q_j אז נוסיף לדקדוק $q_i \to \sigma \mid \sigma q_j$.
4. אנו טוענים שהתקבל דקדוק ליניארי שמאלי לשפה $L = (L^R)^R$.

מעבר בין דקדוק ליניארי שמאלי לימני עובד אותו הדבר. ל-L^R נבנה דקדוק ליניארי שמאלי ואז נהפוך את הכללים באותו האופן.

דקדוק חסר הקשר

הגדרה (דקדוק ח"ה): $G = (T, V, P, S)$ הוא חסר הקשר (ח"ה) אם כל כלל גזירה ב-P הוא מהצורה $A \to \alpha$ כאשר $A \in V$ וגם $\alpha \in (V \cup T)^*$.

הגדרה (שפה ח"ה): נאמר ששפה $L \subseteq \Sigma^*$ היא ח"ה אם קיים דקדוק ח"ה G כך ש: $L = L(G)$. (נובע מכאן שגם שפות רגולריות הן חסרות הקשר, אך **לא להפך**).

- אם מבקשים להוכיח ששפה היא ח"ה אז בונים לה דקדוק ח"ה.

עץ גזירה

הגדרה (עץ גזירה): עץ גזירה עבור דקדוק ח"ה G הינו עץ **סופי**, סדור המקיים:

1. כל צומת מסומנת בסימן מתוך $V \cup T \cup \{\varepsilon\}$.
2. השורש יסומן בדר"כ ב-S (למעט תתי עצים).
3. הסימון של כל צומת פנימי שאינו עלה הוא משתנה מ-V.
4. צומת המסומנת ב-ε הוא בן יחיד.
5. אם צומת פנימית מסומן ב-A ולבניו יש סימונים $X_1, X_2, ..., X_t$ (בסדר הזה!) אזי $(A \to X_1 X_2 ... X_t) \in P$.

לכל מילה שניתן לגזור אותה מהדקדוק יש עץ גזירה.

הגדרה (חזית העץ): המחרוזת המתקבלת ע"י שרשור כל העלים משמאל לימין היא חזית עץ גזירה. אם החזית של העץ מורכבת ממילה טרמינלית, אז העץ הוא עץ גזירה מלא.

טענה:

α היא מילת חזית עץ הגזירה עם שורש $A \leftrightarrow A \Rightarrow^* \alpha$. (הוכח בהרצאה רק כיוון $\leftarrow$).

מסקנה: קבוצת מילות החזית של כל עצי הגזירה המלאים בדקדוק G היא $L(G)$.

טענה (ללא הוכחה): יהי T_1 תת-עץ גזירה של T_2 ותהיינה α_1, α_2 מילות החזית של T_1, T_2 בהתאמה. אזי α_1 תת מילה של α_2.

תבנית פסוקית: זה ביטוי שיש בו גם משתנים וגם אותיות. אם אנו עוצרים באמצע הגזירה ולא ממשיכים עד הסוף אז חזית העץ יכולה להיות ביטוי הכולל גם משתנים.

תת עץ גזירה הוא חלק מהעץ המכיל צומת כלשהו ואת כל צאצאיו.

הגדרה: דקדוק G ייקרא "**רב משמעי**" אם קיימת מילה $w \in L(G)$ עבורה קיימים לפחות שני עצי גזירה. (נרצה להימנע ממצב בו למילה כלשהי יש כמה עצי גזירה שונים).

- דקדוק ייקרא **דו משמעי** אם יש בו מילה שניתן לקבל אותה ב 2 דרכים שונות.
- דקדוק ייקרא **חד משמעי** אם לכל מילה בו יש עץ גזירה יחיד (דרך אחת לגזור אותה).

הגדרה (גזירה שמאלית יחידה): נתונה גזירה $\alpha \to \beta$ כאשר $\alpha, \beta \in (V \cup T)^*$. כלומר, קיימות תבניות פסוקיות $\delta_1, \delta_2, \gamma \in (V \cup T)^*$ ו-$A \in V$ כך ש: $A \to \gamma$, $\alpha = \delta_1 A \delta_2$ וגם $\beta = \delta_1 \gamma \delta_2$.

- הגזירה $\alpha \to \beta$ תקרא "גזירה שמאלית ביותר" $\leftrightarrow$ $\delta_1 \in T^*$.
- הגזירה $\alpha \to \beta$ תקרא "גזירה ימנית ביותר" $\leftrightarrow$ $\delta_2 \in T^*$.
- סדרת הגזירות $\alpha \Rightarrow^* \beta$ תקרא "שמאלית/ימנית ביותר" $\leftrightarrow$ **כל** גזירה בה היא שמאלית/ימנית ביותר.
- כדי לקבל גזירה ימנית/שמאלית ביותר, בכל שלב נפעיל כלל גזירה על המשתנה הימני/שמאלי ביותר.
- כדי לדבר על המילים הנגזרות בדקדוק בצורה חד-משמעית, נשתמש בגזירה שמאלית/ימנית ביותר.
- אם נבצע בדקדוק תמיד גזירה שמאלית/ימנית ביותר אז רוב הבעיה של "דו משמעות" תעלם כי אנחנו יודעים מהו הסדר שבו גוזרים.

משפט: יהי דח"ה G. אזי לכל עץ גזירה ב-G מתאימה גזירה שמאלית ביותר אחת.

פישוט דקדוקים

סדר פעולות 4 אלגוריתמים (**א.י.ט.י**) (מוצג כאן לפי הסדר):

1. ביטול כללי **א**פסילון: $A \to \varepsilon$. (פרט ל $S \to \varepsilon$ במידה והמילה הריקה בשפה)
2. ביטול כללי **י**חידה: $A \to B$.
3. ביטול משתנים לא **ט**רמינליים: משתנים שלא ניתן להגיע מהם למילה סופית.
4. ביטול משתנים לא **י**שיגים: מי שלא ניתן להגיע אליו מ S.

הגדרה: משתנה A ייקרא טרמינלי אם הוא גוזר מילה טרמינלית. כלומר אם $A \Rightarrow^* w$ עבור $w \in T^*$ כלשהי.

הגדרה: משתנה A נקרא **שימושי** אם הוא מופיע בסדרת גזירה המייצרת מילה בדקדוק. אחרת A נקרא **מיותר**.

הגדרה: כלל ε הוא כלל גזירה מהצורה $A \to \varepsilon$. משתנה A נקרא **אפיס** אם $A \Rightarrow^* \varepsilon$.

ביטול כללי אפסילון:

1. נמצא את קבוצת המשתנים האפיסים N.
2. כל כלל גזירה מהצורה $A \to \varepsilon$ יימחק (כולל $S \to \varepsilon$).
3. עבור כל כלל גזירה $A \to X_1 X_2 ... X_k \in P$, $k > 0$, נכניס ל-P' את כלל הגזירה מהצורה $A \to \alpha_1 \alpha_2 ... \alpha_k$ כך ש:
   - i. אם $X_i \notin N$ אזי $\alpha_i = X_i$.
   - ii. אם $X_i \in N$ אזי $\alpha_i \in \{\varepsilon, X_i\}$ פרט ל: $\alpha_1 .... \alpha_k$.
   - iii. כלומר,בהינתן t משתנים אפיסים בכלל הגזירה, ייכנסו ל-P' כל תתי הקבוצות האפשריות מעליהם (משתנה מופיע/לא מופיע , פרט לקבוצה בה כולם ε).

ביטול כללי יחידה:

1. סלק כללי ε (אם S אפיס, הוסף בסוף את $S \to \varepsilon$).
2. מוצאים את כל זוגות המשתנים A, B כך ש: $A \Rightarrow^* B$. ואז נמחק את כללי היחידה.
3. עבור כל זוג כזה, נצרף את כל הכללים שיש ל-B לכללים של A.

ביטול משתנים לא טרמינליים:

1. מוצאים את המשתנים שגוזרים ישירות סימנים שהם טרמינלים ומכניסים את כולם לקבוצה.
2. בכל איטרציה נוספת, מוסיפים לקבוצה את המשתנים שגוזרים תבניות המכילות טרמינלים + משתנים הנמצאים כבר בקבוצה שבנינו.
3. האלגוריתם מסתיים כאשר לא נוספו משתנים חדשים לקבוצה.
4. בסיום האלגוריתם, כל המשתנים שלא נכנסו לקבוצה הם משתנים לא טרמינליים ומוחקים אותם ואת כל הכללים המכילים אותם.

ביטול משתנים לא ישיגים:

1. נקבע $T = \emptyset$, $V' = \{S\}$.
2. לכל כלל $A \in V'$ מהצורה $A \to \alpha$ צרף את כל המשתנים הרשומים ב-α ל- V', ואת כל הטרמינלים (אותיות, אם יש מחרוזת - הכנס כל תו בנפרד) הרשומים ב-α ל-T.
3. חזור על שלב (2) עד שעברת על כל משתנה מ-V' ונעצור כשלא נוסף משתנה חדש.
4. נגדיר P' כללים מ-P המכילים רק משתנים הניתנים להשגה (רק כללים של משתני V').

הצורה הנורמלית של חומסקי

כל כללי הגזירה הם מהצורה $A \to BC$ או $A \to \sigma$

משפט: כל שפה חסרת הקשר (ללא ε) ניתנת לייצור ע"י דקדוק בצורה נורמלית.

אלגוריתם מעבר לצורה נורמלית של חומסקי:

נתרגם את G המייצר את L לצורה נורמלית של חומסקי:

1. סילוק כללי ε (כלומר, כללים מהצורה $A \to \varepsilon$) וסילוק משתנים שלא ניתנים להשגה.
2. סילוק כללי יחידה $(A \to B)$. (זה לא מייצר כללי ε).
3. תרגום כללים מהצורה $A \to X_1 X_2 ... X_k$ (כאשר $k > 2$) לצורה חוקית:
   - a. אם $X_i = \sigma$ אזי נוסיף משתנה חדש C_σ ואת הכלל $C_\sigma \to \sigma$, ונחליף את X_i ב-C_σ. נבצע את זה לכל כלל גזירה ב-P ולכל טרמינל בו.
   - b. לכל $A \to B_1 ... B_k$ עבור $k \geq 3$: נוסיף משתנים $D_1, ..., D_{k-2}$ ונחליף את הגזירה בקבוצת הגזירות כך: $A \to B_1 D_1,\ D_1 \to B_2 D_2, \ . \ . \ . \ ,\ D_{k-3} \to B_{k-2} D_{k-2}, \ . \ . \ .$

למת הניפוח לשפות חסרות הקשר

אם L שפה חסרת הקשר אזי **קיים** n כך **שלכל** מילה $z \in L$ המקיימת $|z| \geq n$ **קיים** פירוק מהצורה $z = uvwxy$ כאשר:

$$(1) \quad |vwx| \leq n \qquad (2) \quad 1 \leq |v| + |x| \qquad (3) \quad \forall i \geq 0:\ uv^i w x^i y \in L$$

בנוסף, אם G, דקדוק ח"ה מהצורה הנורמלית של חומסקי, יוצר את $L \backslash \{\varepsilon\}$ וגם $|V(G)| = k$ אזי $n \leq 2^k$.

- $L = \{a^*\} \cup \{b^j a^{n^2} | 1 \leq n,\ 0 < j\}$ לא ח"ה אבל מתקבלת בלמה (הוכח במטלה).
- $L = b^* c^* a^* \cup a^* \cdot \{b^j c^j a^j | j \geq 1\}$ לא ח"ה אבל מתקבלת (הוכח בהרצאה).

גובה עץ הגזירה של $z \in L$ הינו לפחות $\log_2(|z|) + 1$. (כי העץ העמוק ביותר שיכול להתקבל הוא עץ בינארי מלא עם $|z|$ עלים לכן הגובה לפחות $\log_2(|z|)$. (1+ לטרמינלים).

יהי $V(G) = k$, $n = 2^k$ ותהי $z \in L$ כך ש: $|z| \geq n$. נקבל שגובה העץ מתקיים:

$$k + 1 = \log_2(2^k) + 1 \leq \log_2(|z|) + 1 \leq h$$

כלומר, מסלול ארוך ביותר בעץ מכיל לפחות $k + 2$ צמתים שהם לפחות $k + 1$ צמתים פנימיים שמסמנים משתנים. היות ויש רק k משתנים אז לפי עקרון שובך היונים קיים במסלול משתנה המופיע פעמיים, כלומר שני צמתים המכילים אותו משתנה.

תבנית הוכחה שהשפה L אינה חסרת הקשר לפי למת הניפוח לשפות ח"ה:

נניח בשלילה שהשפה L היא חסרת הקשר ולכן מקיימת את למת הניפוח לשפות חסרות הקשר. יהי $n \in \mathbb{N}$ הקבוע המובטח מהלמה. נבחר $z = \_\_ \in L$ (המילה z צריכה להיות תלויה ב-n) ומתקיים ש: $|z| = \_\_ \geq n$. יהי $z = uvwxy$ פירוק כלשהו המקיים: (1) $|vwx| \leq n$, (2) $|vx| \geq 1$. נתבונן בפירוקים של z המקיימים את תנאי הלמה ונציג מיקומים אפשריים עבור vwx: ______ . עבור כל פירוק נבחר $i = \_\_$ (לא 1!!!) ומכאן $z' = uv^i w x^i y \notin L$ כי _____ . סתירה לכך ש: L מקיימת את הלמה.

בכל הפירוקים מצאנו i עבורו המילה המנופחת (או מפונצ'רת) לא בשפה ולכן L לא ח"ה.

- עקרון לבחירת המילה z: מילה בשפה שאם ננפח אותה (או נוריד ממנה) אפילו מעט (מתוך n התווים הראשונים), היא כבר לא תהיה בשפה.

סגירות לשפות חסרות הקשר

- כל שפה סופית היא חסרת הקשר וכל שפה רגולרית היא חסרת הקשר.

בהינתן L, L_1, L_2 שפות חסרות הקשר ו-R שפה רגולרית, גם השפות הבאות חסרות הקשר:

$L \cap R$ ח"ה	L^R ח"ה	L^n ח"ה

איחוד: $L_1 \cup L_2$ ח"ה. - הוכחה (לא פורמלית): בצעד הגזירה הראשון נבחר האם אנחנו רוצים מילה מ-L_1 או מילה מ-L_2.

שרשור: $L_1 \cdot L_2$ ח"ה - הוכחה (לא פורמלית): בצעד הגזירה הראשון נגזור למשתנה ההתחלתי של L_1 וגם למשתנה של L_2 וכך נקבל מילה מ-L_1 משורשרת עם מילה מ-L_2.

איטרציה: L^* ח"ה - הוכחה (לא פורמלית): בצעד הגזירה הראשון נאפשר לקבל חזרה את S כדי לגזור מילה נוספת מ-L_1. וכשיימאס לנו נבחר ב-ε ונעצור. (אפשר לבחור בו מההתחלה).

שפות ח"ה לא סגורות לתכונות: חיתוך $L_1 \cap L_2$ (לא בהכרח), משלים $\bar{L}$ והפרש.

דוגמה: תהי שפה $L = \{d^*\} \cup \{a^n b^n c^n : n \geq 1\}$. נניח בשלילה כי L חסרת הקשר ותהי $R = \{a^* b^* c^*\}$. נשקול את השפה: $L_\cap = L \cap R = \{a^n b^n c^n : n \geq 1\}$.

לפי תכונת סגור "חיתוך ח"ה עם שפה רגולרית" ציפינו לקבל שפה חסרת הקשר, אך קיבלנו שפה $L_\cap$ שאינה חסרת הקשר! (הוכח בהרצאה) ולכן השפה L לא חסרת הקשר.

אוטומט מחסנית אי דטרמיניסטית (א"מ)

אוטומט מחסנית הוא שביעייה: $M = (Q, \Sigma, q_0, \delta, F, \Gamma, \perp)$.

- Γ - א"ב המחסנית (איזה תווים ניתן להכניס למחסנית).
- δ - פונקצית המעברים: $\delta: Q \times (\Sigma, \varepsilon) \times \Gamma \to 2^{Q \times \Gamma^*}$. הפונקצייה קוראת מצב, אות קלט (או אפסילון) ואת ראש המחסנית ופולטת זוגות של מצב + "מה לכתוב בראש". זה בזוגות מאחר שהאוטומט מחסנית יהיה אי-דטרמיניסטי.
- $\perp$ - תו תחתית המחסנית.
- F - קבוצת מצבים מקבלים (שייך רק במודל קבלה ע"י מצבים מקבלים)

ישנם 2 מודלים (שקולים) שעל פיהם מילה מתקבלת באוטומט:

1. **קבלה ע"י מצבים מקבלים** $L_f(M)$- כמו באוטומט הרגיל.
   $$L_f(M) = \{w \in \Sigma^* | \exists p \in F, \gamma \in \Gamma^* : (q_0, w, \perp) \vdash^* (p, \varepsilon, \gamma)\}$$
2. **קבלה ע"י ריקון** $L_\varepsilon(M)$ - אין מצבים מקבלים אלא אם בסוף קריאת המילה המחסנית ריקה לגמרי (גם לא תו תחתית) אז המילה מתקבלת ואחרת היא לא מתקבלת.
   $$L_\varepsilon(M) = \{w \in \Sigma^* | (q_0, w, \perp) \vdash^* (p, \varepsilon, \varepsilon)\}$$

- אוטומט מחסנית **דטרמיניסטית** חלשה מאוטומט מחסנית אי-דטרמיניסטית.
- דקדוק ח"ה שקול לאוטומט מחסנית. אוטומט סופי עם זיכרון אחד של מחסנית.

הגדרה (תיאור רגעי): עבור אוטומט מחסנית M, תיאור רגעי (ID) הוא השלשה: (q, w, γ) כאשר: $q \in Q$ מצב הבקרה הנוכחי, $w \in \Sigma^*$ יתרת הקלט ו-$\gamma \in \Gamma^*$ תוכן המחסנית.

- לדוגמא: עבור הקלט $aaabbb$ לאחר שני צעדים יהיה: $(q_0, abbb, AA)$.

הגדרה (תיאור רגעי עוקב): נאמר כי $ID_2 = (p, w, \beta\gamma)$ עוקב ל-$ID_1 = (q, \sigma w, Z\gamma)$ אם ורק אם $(p, \beta) \in \delta(q, \sigma, Z)$.

- במילים: אם ממצב q ע"י קריאת הקלט σ וראש המחסנית Z ניתן להגיע למצב p ולרשום במחסנית את β.

סימון: $\vdash^*$ יתאר סדרת צעדי חישוב. כלומר $(q, w, \gamma) \vdash^* (p, x, \beta)$ אמ"מ קיימת סדרת תיאורים רגעיים עוקבים המתחילה ב-(q, w, γ) ומסתיימת ב-(p, x, β).

אוטומט מחסנית (דטרמיניסטית)

נאמר כי אוטומט מחסנית דטרמיניסטי אם הוא מקיים את התכונות הבאות:

- $\forall q \in Q,\ \sigma \in \Sigma,\ Z \in \Gamma: |\delta(q, \sigma, Z)| \leq 1$
- $\forall q \in Q,\ Z \in \Gamma$ אם $\delta(q, \varepsilon, Z) \neq \emptyset$ אזי $\forall \sigma \in \Sigma$ מתקיים: $\delta(q, \sigma, Z) = \emptyset$.

כלומר: מצב מצב, לכל אות קלט ולכל ראש מחסנית יש לכל היותר אפשרות אחת להתקדם (אם ראש המחסנית שונה, זו אפשרות חדשה). בנוסף, אם קיים צעד אפסילון ממצב כלשהו, אזי לא ניתן גם לקרוא אות במצב זה (אלא אם ראש המחסנית שונה).

שקילות אוטומט מחסנית ודקדוק חסר הקשר

מדקדוק לאוטומט מחסנית: לכל שפה ח"ה L קיים M כך ש: $L = L_\varepsilon(M)$.

רעיון ההוכחה: לכל שפה L ח"ה קיים דקדוק ח"ה G שיוצר אותה (לפי הגדרה).

נבנה א"מ M שיחקה גזירה שמאלית ביותר בדקדוק.

התבנית הפסוקית הנוכחית תיוצג ע"י שרשור רישא הקלט שהאוטומט קרא, יחד עם תוכן המחסנית (ב-ID השארנו את מה שנותר לקרוא ואילו כאן יעניין אותנו מה שכבר קראנו).

המטרה: לסיים את מילת הקלט ואת תוכנית המחסנית בבת אחת (L_ε).

אלגוריתם לאוטומט מחסנית:

1. אתחל מחסנית בסימן יחיד S.
2. אם בראש המחסנית מופיע המשתנה הדקדוקי $A \in V$, אז בחר באופן אי דטרמיניסטי באחד מכללי הגזירה $A \to \alpha$ והחלף את ראש המחסנית ב-α (מסע ε).
3. אם בראש המחסנית מופיע b וגם אות הקלט הבאה היא b, קרא אות זו והוצא את b מראש החסנית.
4. חזור לשלב (2).

שימו לב: אין כאן יצירת מצבים. כלומר, נאשר בכל הריצה ב-q_0.

טענה: $\forall x \in T^*$ ולכל $\alpha \in (V \cup T)^*$ **שאינה** מתחילה בטרמינל, $S \Rightarrow^* x\alpha$ בגזירה שמאלית ביותר $\leftrightarrow$ $(q_0, x, S) \vdash (q_0, \varepsilon, \alpha)$. (הוכח בהרצאה באינדוקציה).

מאוטומט מחסנית לדקדוק: בהינתן א"מ M נוכיח כי השפה $L_\varepsilon(M)$ ח"ה ע"י בניית דקדוק ח"ה מתאים. יהי $M = (Q, \Sigma, \Gamma, \delta, q_0, \perp, \emptyset)$. גזירות הדקדוק יבצעו סימולציה של החישובים המקבלים של M. משתני הדקדוק יהיו שלישיות מהצורה $[p, A, q]$ לכל $p, q \in Q,\ A \in \Gamma$ והמשתנה ההתחלתי S. שלשה כזו תגזור מילה טרמינלית $\leftrightarrow$ קריאת x ע"י M עשויה לגרום למחיקת A מראש המחסנית ומעבר ממצב p למצב q.

משפט: לכל א"מ M השפה $L_\varepsilon(M)$ הינה ח"ה. (הוכח בהרצאה).

בעיות הכרעה לאוטומטים

בעיית השייכות: בהינתן מילה w ואוטומט A, האם $w \in L(A)$?

פתרון: נפעיל את A על w ונבדוק את מצבו בסיום קריאת המילה.

בעיית הריקנות: בהינתן אוטומט A, האם $L(A) = \emptyset$?

פתרון: שני פתרונות: פתרון (1) מתבסס על משפט שקובע תנאי הכרחי ומספיק לריקנות. פתרון (2) ויעיל יותר בהרבה יתבסס על אלגוריתם בגרפים.

משפט 1 (הוכח בהרצאה): A אס"ד. $L(A) \neq \emptyset \Leftrightarrow \exists z \in L(A)$ כך ש: $|z| < |Q|$.

פתרון (1): נעבור על כל המילים שאורכן קטן מ-$|Q|$ בסדר כלשהו, ונפעיל על כל אחת מהם את "אלגוריתם השייכות". אם המילה שהופעלה שייכת ל-$L(A)$ אז נעצור ונשיב כי השפה $L(A)$ אינה ריקה. אחרת, נעצור כאשר עברנו את כל המילים ונשיב כי $L(A)$ ריקה.

סיבוכיות (1): מספר המילים באורך i הוא $|\Sigma|^i$, האלגוריתם יבדוק $\frac{|\Sigma|^{|Q|}-1}{|\Sigma|-1}$ מילים.

פתרון (2): נבדוק האם קיים מסלול מ-q_0 למצב מקבל, במידה ולא, השפה $L(A)$ ריקה.

ניתן למצוא ביעילות את המצבים הניתנים להשגה באמצעות BFS על האוטומט.

סיבוכיות (2): סיבוכיות של BFS הוא $O(|Q| \cdot |\Sigma|)$.

בעיית הסופיות: האם שפת אוטומט נתון סופית או אינסופית.

משפט 1: השפה המתקבלת ע"י אס"ד A בעל n מצבים היא אינסופית אם"ם קיימת מילה $z \in L(A)$ כך ש: $n \leq |z| < 2n$. (כלומר, מספיק וקיימת מילה אחת באורך לפחות $|Q|$ בשביל לקבל שהשפה אינסופית, ומספיק לנו לבדוק עד $2n$ - מעגל ראשון באוטומט).

פתרון (1): נעבור על כל המילים $n \leq |w| < 2n$ ואם נמצא מילה שהאוטומט מקבל אזי השפה אינסופית. אחרת, השפה סופית. (לפי משפט 1).

סיבוכיות (1): במקרה הגרוע יש לבדוק מספר מעריכי של מילים: $|\Sigma|^n \cdot \frac{|\Sigma|^n-1}{|\Sigma|-1}$.

משפט 2: יהי A אס"ד שכל מצביו ניתנים להשגה. יהי $G(A)$ הגרף המתאים לו. אזי A אינסופית $\Leftrightarrow$ קיים ב-$G(A)$ מעגל המכיל מצב שניתן להגיע ממנו למצב מקבל.

פתרון (2): **א**. מחק מ-$G(A)$ כל מצב שאינו ישיג, וגם את כל המצבים שאין מסלול מהם למצב מקבל. **ב**. בדוק האם קיים מעגל בגרף שהתקבל.

סיבוכיות (2): מעבר בגרף לבדיקת מעגל באמצעות BFS יהיה $O(|Q| \cdot |\Sigma|)$.

בעיית השקילות: בהינתן שני אס"דים A_1, A_2, האם $L(A_1) = L(A_2)$?

פתרון: מתוך הבניות של אוטומט לחיתוך ומשלים של שפות נוכל לבנות אוטומט המקבל את שפת ההפרש הסימטרי: $L = L_1 \Delta L_2 = (L_1 \backslash L_2) \cup (L_2 \backslash L_1)$.

מתקיים: $L_1 = L_2 \leftrightarrow L = \emptyset$. נפעיל את האלגוריתם היעיל ל"בעיית הריקנות" לשפה L ובמידה וחזר לנו שהשפה L ריקה אז האוטומטים שקולים, אחרת הם לא שקולים.

בעיות הכרעה לשפות חסרות הקשר

בעיית השייכות: בהינתן דקדוק G ומילה w, האם $w \in L(G)$?

פתרון (1): אם $w = \varepsilon$ פשוט נבדוק האם S אפיס. אם $w \neq \varepsilon$ אז נעזר בחומסקי: נמיר את הדקדוק G לצורה הנורמלית של חומסקי. אם $w \in L(G)$ קיימת סדרת גזירה באורך בדיוק $2|w| - 1$ שגוזרת אותה (בכל צעד מורידים משתנה ומוסיפים 2 ועוד $|w|$ צעדים לגזור טרמינלים). יש כמות סופית של סדרות גזירה באורך זה, אם אחת מהם גוזרת את w אז היא בשפה, אחרת היא לא.

סיבוכיות (1): רץ בזמן אקספוננציאלי כי צריך לעבור על המון עצי גזירה.

פתרון (2) אלגוריתם CYK: יש בדף נוסחאות הסבר עליו בעמוד הבא.

סיבוכיות (2): $O(|w|^3)$.

פתרון (3) עבור א"מ דטרמיניסטי בלבד.

סיבוכיות (3): זמן ליניארי.

בעיית הריקנות: בהינתן דקדוק G, האם $L(G)$ ריקה?

פתרון (וגם הוכחה): נריץ את האלגוריתם לסילוק משתנים לא טרמינליים ונבדוק האם S משתנה טרמינלי: S משתנה טרמינלי $\leftrightarrow L(G) \neq \emptyset$.

משפטים

משפט דיריכלה: אם $(a, b) = 1$ אזי קיימים אינסוף ראשוניים מהצורה $an + b$.

מספר פריק: כל מספר שלם מ-1 הוא או ראשוני או פריק. המספר פריק אם ניתן לכתוב אותו כמכפלה של שני שלמים גדולים מ-1.

מטלות

תרגיל: יהי הדקדוק הבא: $G = (\{A, S\}, \{a, b\}, S, P)$ כאשר כללי הדקדוק הם: $A \to aAb|ab$.$S \to aAb|Sb$. **(א)** מהי שפת הדקדוק $L(G)$? **(ב)** את השפה שרשמתם לעיל סמנו כ-L והוכיחו כי $L(G) = L$ באינדוקציה (שני הכיוונים). **פתרון (א)**:
$L(G) = \{a^n b^m | 2 \le n \le m\}$. ניתן להוסיף עוד ועוד b עד שבוחרים בכלל $S \to aAb$ ואז הכמות היא זהה בפורמט של קודם a ואז b. אבל לפחות 2 תווים מכל אחד. **פתרון (ב)**:
כיוון 1: נוכיח כי $L \subseteq L(G)$. **נוכיח טענת עזר**: $\{a^n b^n | n \ge 1\} \subseteq L(A)$. הוכחת טענת העזר: באינדוקציה על אורך המילה: **בסיס**: $w = ab$. ונקבל: $A \to ab$. **צעד**: נניח נכונות לכל מילה עד אורך n ונוכיח עבור מילים באורך $n + 1$. תהי $w \in \{a^n b^n | n \ge 1\}$ מילה כזו. מכאן: $w = a^m b^m = aa^{m-1}b^{m-1}b = axb$ עבור $m \ge 2$ כאשר $x \in \{a^n b^n | n \ge 1\}$ ולכן לפי הנחת האינדוקציה: $A \to^* x$ ולכן: $A \to aAb \to^* aa^{m-1}b^{m-1}b = w$. נוכיח את **הטענה הראשית** באינדוקציה על אורך המילה: **בסיס**: $aabb$. נקבל: $S \to aAb \to aabb$. **צעד**: נניח נכונות עבור מילים עד אורך n ונוכיח עבור אורך $n + 1$. תהי $w \in L$ מילה כזו. מכאן: $w = a^n b^m$ כך ש: $2 \le n \le m$. אם $n < m$ אז ניקח: $w = a^n b^{m-1} b$. מתקיים: $a^n b^{m-1} \in L$ ולפי הנחת האינדוקציה מתקיים: $S \to^* a^n b^{m-1}$. ולכן: $S \to Sb \to^* a^n b^{m-1} b = w$. אם $n = m$ אז ניקח: $w = aa^{n-1}b^{n-1}b$ כאשר $1 \le n$. מכאן לפי טענת העזר נקבל: $S \to aAb \to aa^{n-1}b^{n-1}b = w$. **כיוון 2**: נוכיח $L(G) \subseteq L$. נוכיח **טענת עזר**: $L(A) \subseteq \{a^n b^n | n \ge 1\}$. הוכחת טענת העזר: באינדוקציה על אורך הגזירה: **בסיס**: $A \to ab$ ואכן: $ab \in \{a^n b^n | n \ge 1\}$. **צעד**: נניח נכונות לכל מילה הנגזרת עד n צעדים ונוכיח עבור מילים הנגזרות לאחר $n + 1$ צעדים. תהי w מילה כזו כך ש $A \to^{n+1} w$. מכאן לפי הגדרת כללי הגזירה, בצעד הראשון השתמשנו בהכרח בכלל: $A \to aAb$. לפי הנחת האינדוקציה, בשאר n צעדי הגזירה נקבל: $A \to^n x \in \{a^n b^n | n \ge 1\}$. ומכאן: $w = axb = aa^k b^k b \in \{a^n b^n | n \ge 1\}$ כנדרש. נוכיח את **הטענה הראשית** באינדוקציה על אורך הגזירה: **בסיס**: $S \to aAb \to aabb$ ואכן: $aabb \in L$.
צעד: נניח נכונות לכל מילה שנגזרת לאחר לכל היותר n צעדי גזירה ונוכיח עבור מילים הנגזרות לאחר $n + 1$ צעדי גזירה. תהי w מילה כזו. מכאן: $S \to^{n+1} w$. **מקרה א**: $S \to Sb \to^n xb = w$ אז לפי הנחת האינדוקציה: $x \in L$ ולכן: $x = a^n b^m$ כאשר $2 \le n \le m$. ולכן $w = a^n b^{m+1} \in L$. **מקרה ב**: $S \to aAb \to axb = w$. אז לפי טענת העזר נקבל ש $x \in \{a^n b^n | n \ge 1\}$ ולכן: $w = aa^n b^n b \in L$.

תרגיל: הוכיחו או הפריכו: תהי השפה $L = \{w \in \Sigma^* : |w| \ge 1000\}$ ותהי P שפה לא רגולרית. אזי בהכרח $L \cup P$ רגולרית. **פתרון**: הוכחה: נעזר בדה-מורגן: $L \cup P = \overline{\bar{L} \cap \bar{P}}$ נשים לב כי $\bar{L} = \{w \in \Sigma^* : |w| < 1000\}$ כלומר שפה סופית. לכן $\bar{L} \cap \bar{P}$ גם שפה סופית ומכאן שהיא רגולרית. מסגירות למשלים, גם $L \cup P$ רגולרית. **תרגיל**: השפות הלא רגולריות סגורות לחיתוך. **פתרון**: הפרכה: $L_1 = \{0^n 1^n | 1 \le n\}$ ו-$L_2 = \{1^n 0^n | 1 \le n\}$ אזי $L_1 \cap L_2 = \emptyset$. כלומר, חיתוך של שתי שפות לא רגולריות יצא שפה רגולרית - סתירה.

תרגיל: יהיו $L_1, L_2, ..., L_{2n}$ שפות רגולריות. בנו אוטומט לשפה הבאה:
$L = \{w_1 ... w_{2n} | \forall i \in [1, 2n]:\ w_i \in L_1\}$. הנחיה קלה: יש לתאר בנייה פורמלית של האוטומט. **פתרון**: השפות $L_1, ..., L_{2n}$ רגולריות ולכן קיים לכל אחת מהן $NFA - \varepsilon$ בעל מצב מקבל יחיד, נקרא להם $E_1 ... E_{2n}$ בהתאמה. נסמן $E_1 = (Q_1, \Sigma, \delta_1, q_{01}, q_{f1})$ וכן הלאה. נעזר בהם בשביל לבנות $NFA - \varepsilon$ עבור השפה L בצורה הבאה:
$E(L) = (\{Q_1\} \cup \{Q_2\} \cup ... \cup \{Q_{2n}\} \cup \{q_{0new}, q_f\}, \Sigma, \delta', q_f)$.
$\delta'(q_{0new}, \varepsilon) = q_{01}\ (first\ state\ of\ Q_1)$. כל מצבי q_{fi} שהיו מקבלים באוטומטים $E_1 ... E_{2n}$ אינם מקבלים יותר. עבור $\forall i \in [1, 2n],\ \forall q \in \{Q_i\},\ \forall \sigma \in \Sigma$ מתקיים:
$\delta'(q, \sigma) = \delta_i(q, \sigma)$. כלומר, האוטומט החדש מתנהג זהה לאוטומטים המקוריים בתוכם.
$\forall i \in [1, 2n - 1]: \delta'(q_{fi}, \varepsilon) = q_{0(i+1)}$. כלומר, חידשנו מסע ε למצב הראשון של האוטומט הבא, מסוף האוטומט הנוכחי. $\delta'(q_{f2n}, \varepsilon) = q_f$. ובסוף קריאת המילה האחרונה, הלכנו למצב המקבל היחיד החדש במסע ε. למעשה, הכללנו את הבנייה שראינו בהרצאה של תכונות סגור "שרשור" כך שהיא תפעל על $2n$ שפות.

תרגיל: תהי השפה הבאה: $L_n = \left\{w \in \{0, 1, 2\}^*\ |\ the\ n^{th}\ from\ last\ char\ is\ 1\right\}$
בהינתן n טבעי כלשהו, בנו אס"ד לשפה. **פתרון**: נבנה את האס"ד הבא:
$A = (Q, \Sigma, q_0, \delta, F)$. $Q = \{q_{a_1a_2...a_n} | \forall i \in [n]: a_i \in \{0, 1, 2\}\}$.
$F = \{q_{1a_2a_3..a_n} | \forall i \in [2..n]: a_i \in \{0, 1, 2\}\}$. $q_0 = \{q_{a_1a_2...a_n} | \forall i \in [n]: a_i = 0\}$
מעברים: $\forall \sigma \in \{0, 1, 2\}, \forall q_{a_1a_2...a_n} \in Q: \delta(q_{a_1a_2..a_n}, \sigma) = q_{a_2a_3...a_n\sigma}$.

תרגיל: נגדיר מודל סופי דטרמיניסטי חדש $starting - set = \left(Q, \Sigma, \tilde{q}_0,\ \delta,\ F\right)$ כאשר $\tilde{q}_0 \subseteq Q$ (כלומר, יש קבוצה של מצבים התחלתיים אפשריים). נגדיר כי מילה מתקבלת באוטומט: $L(startingset) = \left\{w \in \Sigma^{\star}\ |\ \exists q \in \tilde{q}_0\ \ s.t.\ \ \hat{\delta}(q, w) \in F\right\}$
הוכיחו/הפריכו: מודל זה שקול לאס"ד. (הפרכה: דוגמה נגדית. כלומר, שפה שאחד המודלים יכול לקבל והשני לא. הוכחה: בהינתן אוטומט ממודל אחד לבנות מהשני, ולהפך).
פתרון: המודל אכן שקול לאס"ד. נעזר בשקילות $DFA \equiv \varepsilon NFA$ ונוכיח בצורה הבאה:
בהינתן $A = (Q, \Sigma, q_0, \delta, F)$ נבנה אס"ד מהמודל החדש: $A' = (Q, \Sigma, \{q_0\}, \delta, F)$ וקל לראות שהאוטומט מהמודל החדש מקבל אותה שפה בדיוק. עבור $A' = (Q, \Sigma, \tilde{q}_0, \delta, F)$ נבנה εNFA שקול $E = (Q \cup \{q_{0new}\}, \Sigma, \delta', F)$. נותר להגדיר את δ': לכל $\sigma \in \Sigma$ ו-$q \in Q$ מתקיים: $\delta'(q, \sigma) = \delta(q, \sigma)$. (פונקציית המעברים של εNFA מתנהגת כמו פונקציית המעברים של A' לכל המצבים שאינם q_{0new}). $\delta'(q_{0new}, \varepsilon) = \{\tilde{q}_0\}$. (מהמצב ההתחלתי החדש של האוטומט ניתן לגשת במסע ε לכל אחד מהמצבים ההתחלתיים של האוטומט A'). היות והשינוי היחיד הוא צעד בודד ללא קריאת קלט מהמצב הראשון החדש לקבוצה המצבים הראשונים במודל המקורי, השפות שקולות.

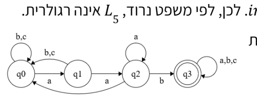
תרגיל: העזרו בשובך היונים והוכיחו כי השפה הבאה לא רגולרית: $L = \{a^n b^m : n \ne m\}$.
פתרון: נניח בשלילה כי השפה רגולרית ויהי A האס"ד שלה. נסמן $|Q| = n$. נשקול את המילים: $\varepsilon, a, a^2, ..., a^n$ יש כאן $n + 1$ מילים אבל באוטומט רק n מצבים ולפי העקרון קיימים $i \ne j$ המקיימים: $\delta(q_0, a^i) = q_u = \delta(q_0, a^j)$. נשרשר b^i ונקבל:
$\delta(q_0, a^i b^i) = \delta(\delta(q_0, a^i), b^i) = \delta(q_u, b^i) \notin F$ שכן $a^i b^i \notin L$. לעומת זאת:
$\delta(q_0, a^j b^i) = \delta(\delta(q_0, a^j), b^i) = \delta(q_u, b^i) \notin F$ שכן $a^j b^i \in L$ וזה סתירה!

מטלות

תרגיל: במטלה הוצג גרסה מוכללת ללמת הניפוח לרגולריות (עם $tuv^i wg$). היעזרו בגרסה זו על מנת להוכיח כי: $L = \{b^j a^{k^2} c^r | 1 \le j, k, r\}$ אינה רגולרית. **פתרון**: נניח בשלילה כי L רגולרית ולכן היא מקיימת את למת הניפוח המוכללת. תהי $z = a^n b^{n^2} c^n \in L$ המקיימת $n \le |z| = n^2 + 2n$. נסתכל על הפירוק: $t = a^n, s = b^{n^2}, g = c^n$ המקיים $n \le |s| = n^2$ ולכן קיים פירוק $z = uvw$ העונה על תנאי הלמה. נבחר $i = 2$ ונקבל:
$z^2 = tuv^2wg = a^n b^{n^2+|v|} c^n \notin L$ שכן $n^2 < n^2 + |v| \le n^2 + n < (n + 1)^2$
בסתירה לגרסה המוכללת של למת הניפוח, ולכן L אינה רגולרית.

תרגיל 1: $L = \{a^c | c\ is\ composite\}$. הראו כי הגרסה המוכללת של למת הניפוח לרגולריות לא תורמת להוכחת אי הרגולריות. **פתרון**: בגלל שהשפה מורכבת מתו יחיד בא"ב הקלט אז לא משנה היכן n האותיות שננפח ולכן הגרסה של הלמה עדיין לא עוזרת.
תרגיל 2: לאותו שפה L (composite) הוכיחו כי L אינה רגולרית בעזרת למת הניפוח (הרגיל) ובעזרת משפט דיריכלה: אם $(a, b) = 1$ אזי קיימים אינסוף ראשוניים מהצורה $an + b$.
פתרון : נניח בשלילה כי L רגולרית ויהי n הקבוע המובטח מהלמה. יהי $p > n$ ראשוני כלשהו ונשקול את המילה $z = a^{p^2} \in L$ המקיימת $|z| = p^2 > p > n$ ולכן קיים פירוק $z = uvw$ המקיים $1 \le |v|, |uv| \le n$. היות ו-$n < p$ ו-$1 \le |v| \le n$ אזי בהכרח $(|v|, p) = 1$ ולכן גם $(|v|, p^2) = 1$. לכן לפי משפט דיריכלה, קיימים אינסוף מספרים k עבורם $|v|k + p^2$ ראשוני. נבחר $i = k + 1$ ונקבל: $a^{p^2+(i-1)|v|} = a^{p^2+k|v|} \notin L$ בסתירה ללמה.

תרגיל: הוכיחו/הפריכו: השפות הבאות רגולריות (לפי **נרוד**): $(1)\ L_1 = \{a^p | p\ is\ prime\}$
$(2)\ L_2 = \{a^{n!} | n \in Z^+\}$, $(3)\ L_3 = \{w \in \{a, b, c\}^* | w\ contains\ 'aab'\}$.
פתרון 1: מכיוון ש: L_1 שפה אינסופית, מספיק להראות שכל זוג מילים מהשפה במחלקות שקילות נפרדות. נניח בשלילה כי עבור זוג ראשוניים $p < q$ המילים $a^p,\ a^q$ באותה מחלקת שקילות. מכאן שאם $a^p a^{q-p} = a^q \in L_1$ אז גם $a^q a^{q-p} = a^{2q-p} \in L_1$ ולכן גם $a^{2q-p} a^{q-p} = a^{3q-2p} \in L_1$. באופן כללי, ניתן להוכיח באינדוקציה כי $a^{p+i(q-p)}$ לכל $i \ge 0$. אבל עבור $i = p$ אז $a^{p+p(q-p)} = a^{p(1+q-p)} \in L_1$ כלומר קיבלנו מספר פריק וזו סתירה.
פתרון 2: נגדיר את קבוצת המילים האינסופית $A = \{a^{n!} | 1 \le n\}$. יהיו $i \ne j$, כלומר $a^{i!}, a^{j!}$ שתי מילים כלליות בשפה. נשקול את הסיפא $z = a^{(i+1)!-i!}$. ומתקיים כי:
$a^{i!} z = a^{(i+1)!} \in L_2$ אבל $a^{j!} z = a^{(i+1)!-i!+j!} \notin L_2$ לכל שתי מילים בקבוצה אינסופית קיימת סיפא מפרידה ולכן $index(R_L) = \infty$. לכן, לפי משפט נרוד, L_5 אינה רגולרית.
פתרון 3: נבנה אס"ד לשפה, וממנו נגדיר את מחלקות השקילות של R_L. לכן מחלקות השקילות הן:



$$S_0 = (b + c)^* a((b + c)^* + a^+ c) + (b + c)^*$$
$$S_1 = S_0 a, \quad S_2 = S_0 aa^+, \quad S_3 = S_0 aa^+ b(a + b + c)^*$$
המחלקות זרות ומכסות את כל Σ^*, כל שתי מילים מאותה מחלקה אינן ניתנות להפרדה וכל שתי מילים ממחלקות שונות כן ניתנות להפרדה. $index(R_L) = 4$ ולכן רגולרית.

תרגיל: תהי $L = \{w \in \{0, 1\}^* | \#_1(w) \equiv 0, 5(mod\ 6)\}$. **א.** מהן מחלקות השקילות של R_L? **ב.** ממשו DFA מינימלי הדרוש לשפה זו. **ג.** הוכיחו/הפריכו: קיימת L' (המקיימת $L \ne L'$) כך ש $R_{L'}$ מעדן את R_L. **ד.** הוכיחו/הפריכו: קיימת L'' (המקיימת $L \ne L''$) כך ש R_L מעדן את $R_{L''}$. **ה.** הוכיחו/הפריכו: קיימת אינסוף שפות $\tilde{L}$ עבורן $R_{\tilde{L}}$ מעדן את R_L. **ו.** כיצד תשתנה התשובה של סעיף ב' עבור $L = \{w \in \{0, 1\}^* | \#_0(w) \equiv 1, 4(mod\ 6)\}$?
פתרון: **(א)** מח' השקילות הן: $\forall i \in [0, 5]:\ S_i = \{w \in \{0, 1\}^* | \#_0(w) \equiv i(mod\ 6)\}$
(ב) היות ויש 6 מחלקות שקילות, נדרשות שישה מצבים באוטומט (ציירו את האוטומט במבחן - חלק מפתרון). **(ג)** הטענה נכונה: תהי $L' = \{w \in \{0, 1\}^* | \#_1(w) \equiv 3(mod\ 12)\}$. קל להראות כי ל-$R_{L'}$ יש 12 מח' שקילות וכי כל מחלקה שלו מוכלת במחלקה של R_L. לכן $R_{L'}$ מעדן את R_L כנדרש. **(ד)** הטענה נכונה: תהי $L'' = \Sigma^*$. קל להראות כי ל-$R_{L''}$ יש מחלקת שקילות אחת (ראינו בהרצאה) ולכן כל אחת מהמחלקות של R_L מוכלת במחלקה הבודדת של $R_{L''}$. **(ה)** הטענה נכונה: היות ויש אינסוף טבעיים, יש אינסוף כפולות של 6 ולכן R_L של כל שפה בה נלך על מודולו של $6k$ יעדן את R_L. דוגמה לבנייה כזו: יהי $N = 6k$ עבור k כלשהו. אזי $L_N = \{w \in \{0, 1\}^* | \#_1(w) \equiv 1(mod\ N)\}$ מקיימת כי R_{L_N} מעדן את R_L. **(ו)** כעת מספיקות שלוש מחלקות שקילות שכן $1, 4(mod\ 6) \equiv 1(mod\ 3)$ (חסר בדף נוסחאות ציור של אוטומט, פשוט לצייר 3 מצבים, מכל מצב i ל-i+1 להעביר "1" ומצב 1 הוא המקבל).

תרגיל: בתחילת הוכחת נכונות אלגוריתם $Hopcroft$, הראנו שהיחס $\bar{R}$ הינו יחס שקילות בסוף הריצה. הוכיחו כי היחס מהווה שקילות **בכל סיום איטרציה**. **פתרון**: נוכיח כי היחס $\bar{R}$ הינו יחס שקילות בסיום כל איטרציה של האלגוריתם. נסתמך על האינווריאנטה שהוכחה בהרצאה (בניסוח טיפה אחר): "בשלב ה-i האלגוריתם מכניס ליחס R את הזוג שמופרד על ידי מילה באורך i". מטענה זו נקבל כי בסוף השלב ה-i מתקיים $(p, q) \in \bar{R}$ אמ"מ הן אינן מופרדות ע"י סיפא באורך i ומטה. נותר להראות כי $\bar{R}$ הינו יחס שקילות. היחס $\bar{R}$ הינו רפלקסיבי וסימטרי (היחס מוגדר על זוגות מבלי חשיבות לסדר) ונותר רק להראות שהוא טרנזיטיבי. נניח בשלילה $\bar{R}$ לא טרנזיטיבי ויהיו p, q, r כך שמתקיים: $\bar{R}(p, q)\ \wedge \bar{R}(q, r)\ \wedge$ $\wedge\ \bar{R}(p, r)$. לכן קיימת סיפא z עד אורך i המפרידה בין p, r. כלומר עבור $y \in L(r)$, $x \in L(p)$ מתקיים $xz \in L$ אך $yz \notin L$ (או להפך). נשקול $w \in L(q)$: מה יקרה כשנשרשר את אותה סיפא z? מקרה א': $wz \in L$ ולכן z מפרידה בין r, q. מקרה ב': $wz \notin L$ ולכן z מפרידה בין p, q. סתירה.

תרגיל: הוכיחו או הפריכו: קיים ביטוי רגולרי $r \in R_{0,1}$ המקיים $L[r] = \{1^*\}$ ואין בו שימוש בכלל היצירה $r \to r^*$. הנחיה: היעזרו באינדוקציה מבנית. **פתרון**: נוכיח באינדוקציה שביטוי רגולרי שלא משתמש בכלל יצירה "איטרציה" תמיד ייצר שפה סופית. בסיס $L[\emptyset] = \emptyset,\ L[\varepsilon] = \{\varepsilon\},\ \forall \sigma \in \Sigma:\ L[\sigma] = \{\sigma\}$. נניח שכל ביטוי רגולרי שיש לו עד n צעדי יצירה שאינם איטרציה יוצר שפה אינסופית ונשקול ביטוי רגולרי עם $n + 1$ צעדי יצירה שאינם איטרציה. יהיו r_1, r_2 שני ביטויים רגולריים שלכל אחד מהם n צעדי יצירה. הצעד הנוסף יהיה + או · . עבור + נקבל $L[r_1] \cup L[r_2]$ אך שתיהן מהנחת האינדוקציה ולכן איחוד שפות סופיות נשאר סופי, שכן האיחוד הוא חיבור גדלי השפות פחות החיתוך (הכלה והדחה). עבור · נקבל $L[r_1] \cdot L[r_2]$ אך שתיהן סופיות מהנחת האינדוקציה וראינו בהרצאה הראשונה כי שפת השרשור חסומה במכפלת גדלי השפות המקוריות, כלומר, גם סופית.

(המשך) בעיות הכרעה לשפות חסרות הקשר

בעיית הסופיות: בהינתן דקדוק G, האם $L(G)$ סופית או אינסופית?
משפט: בהינתן דקדוק ח"ה G קיים אלגוריתם המכריע האם $L(G)$ סופית. (הוכח).
הוכחה: נניח בה"כ כי $\varepsilon \notin L(G)$ וכי ב-G אין כללי יחידה, כללי ε וסימנים מיותרים (אם יש כאלה, נוכל להפעיל את האלגוריתמים המסלקים אותם).
נבנה גרף מכוון $D = (V, E)$ בו קבוצת הצמתים V היא קבוצת המשתנים של G והצלעות מוגדרות כך: $E = \{(A, B)\ |\ A \to \alpha B\beta \in P,\ \alpha, \beta \in (V \cup T)^*\}$.
ב-G אין כללי יחידה ולכן $|\alpha\beta| \ge 1$ כלומר $\alpha \ne \varepsilon$ או $\beta \ne \varepsilon$.
טענה: $L(G)$ אינסופית $\Leftrightarrow$ בגרף D קיים מעגל מכוון.
הוכחה (חלקית מההרצאה): (כיוון א'): נניח כי ב-D יש מעגל מכוון, ... , נציג את סדרת הגזירה: $A_n \to \alpha_1 A_2 \beta_1 \to ... \to \alpha_1 ... \alpha_n A_1 \beta_n .... \beta_1$. מאחר ואין בדקדוק סימנים מיותרים (לפי המשפט הקודם) קיימות מילים $v, w, x \in T^*$ כך ש: $\alpha_1 .. \alpha_n \to^* v$, $A_1 \to^* w$, $\beta_n .... \beta_1 \to^* x$ (זה לפי למת הניפוח), קיימת סדרת גזירה: $S \to^* uA_1y$. לסיכום נוכל לרשום את הסדרה: $S \to^* uA_1y \to^* uwy$ או $S \to^* uA_1y \to^* ... \to^* uv^iwx^iy$.
(כיוון ב'): נניח כי $L(G)$ אינסופית ו-n משתנים. ... יש עץ גזירה שאורכו לפחות $n + 1$. מסלול כזה מכיל לפחות $n + 2$ סימנים, כלומר, לפחות $n + 1$ משתנים. לפי שובך היונים, קיים משתנה A_1 המופיע פעמיים. לכן קיימים משתנים $A_2, A_3, ..., A_t$ כך ש: $A_1 \to A_2 \to A_3 \to ... \to A_t \to A_1$ הוא מסלול בעץ הגזירה ולכן נקבל שזה מעגל ב-D.

אלגוריתם CYK (בעיית השייכות לדח"ה):
נניח כי המשתנים בדקדוק הם $R_1 = S,\ R_2, ..., R_r$ והמשתנה ההתחלתי הוא R_1.
נגדיר פונקציה f ש**בודקת האם ניתן לגזור ממשתנה** R_a **תת מחרוזת של** w, ומחזירה בהתאם $T\ or\ F$. הרעיון הוא רקורסיבי: משתנה גוזר תת-מילה, אם הוא גוזר שני משתנים שגוזרים את הרישא והסיפא של תת-המילה. הקלט של הפונקציה הוא שלשה: ℓ אורך תת מילה, j אינדקס תו ראשון של התת מילה ו-a משתנה ממנו גוזרים:
- $f(\ell, j, R_a) = T$ אם $\exists 1 \le k < \ell,\ \exists b, c$ כך ש:
  $R_a \to R_bR_c \wedge f(k, j, R_b) \wedge f(\ell - k,\ j + k + 1,\ R_c)$.
- $f(\ell, j, R_a) = F$ אחרת.

המימוש יהיה בשיטת תכנית דינמי. נאתחל מטריצה בוליאנית P בגודל $n \times n \times r$. נמלא אותה בהתאם לפונקציה. נגדיל בהדרגה את ℓ, האורכים של תתי המילה. בכל איטרציה נתבסס על התוצאות הקודמות שחישבנו. **אתחול** $\ell = 1$: כל המטריצה מאותחלת ב-F, מלבד תתי מחרוזות באורך 1 לפי כללי יחיד של הדקדוק אם $r_v \to w[j]$ אז נמלא $P[1, j, v] = T$. **צעד**: לכל תת מחרוזת באורך $\ell = 2, ..., n$ המתחילה באינדקס $j = 1 ... n - \ell + 1$ יש לבדוק האם ניתן לפצל אותה לשניים לאחר $k = 1 ... \ell - 1$ תווים ולגזור כל חלק בנפרד. כלומר אם $P[k, j, b] \wedge P[\ell - k, j + k, c] \wedge R_a \to R_bR_c$. נעדכן $P[\ell, j, a] = T$. (יש כאן לולאה מקוננת על ℓ, j, k).**סיום**: נחזיר את $P[n, 1, 1]$ כאשר $n = |w|$ ו-1 הוא האינדקס הראשון של w. כך נוכל לדעת האם ניתן לגזור את w מתוך הדקדוק. **זמן ריצה**: $O(n^3)$.

תרגיל: איזה שינוי צריך לבצע לאלגוריתם CYK כך שבהינתן דקדוק ח"ה G ומילה w יקבע האם $w \in Suffix(L(G))$? נמקו. ניתן להניח ש: $w \ne \varepsilon$ וגם G הוא בצורה הנורמלית של חומסקי, ולא מכיל משתנים לא טרמינליים.
פתרון: נרצה שהאלגוריתם יגזור את w ועוד "משהו" אחריה. מכיוון שאין משתנים אפיסים בצורה הנורמלית של חומסקי, מספיק להראות ש: $S \Rightarrow^*\ Aw$ עבור $A \in V$ כלשהו. יש כאן בעצם $n + 1$ תווים, שהראשון הוא משתנה כלשהו (תיגזר ממנו מחרוזת, כי כל המשתנים טרמינליים). כאן, נצטרך להגדיר מערך בגודל $(n + 1) \times (n + 1) \times r$ ולשנות מעט את שלב האתחול. בדומה לקודם, עבור $2 \le j \le n + 1$ אם $r_v \to w[j]$ אז נמלא $P[1, j, v] = T$. יתר התאים יהיו F. אך בנוסף, לכל $v \in [r]$ נמלא $P[1, 1, v] = T$. ההמשך זהה ל-CYK על מילה באורך $n + 1$.

מטלות

תרגיל: תהי L שפה המיוצרת בדקדוק הבא: $G = (V = \{S\}, T = \{a, b\}, S, P)$.
כך ש: $P = \{S \to aSa|bSb|aa|bb\}$. **מצאו את ה-$n$ הקטן ביותר** עבורו הלמה לשפות ח"ה מתקיימת עבור שפה זו. (יש לתת חסם יעיל ל-n ולא למצוא אותו במדויק).
פתרון: צריך קודם להמיר את הדקדוק לצורה הנורמלית של חומסקי, לפי האלגוריתם שנלמד בהרצאה. שלב 1: נוסיף את המשתנים C_a, C_b ל-V.
ונחליף: $S \to C_aSC_a|C_bSC_b|C_aC_a|C_bC_b$. נותר לטפל בשני כללי הגזירה בהם מופיעים שלושה משתנים: הכלל $S \to C_aSC_a$ יהפוך ל: $S \to C_aD,\ D \to SC_a$ והכלל $S \to C_bSC_b$ יהפוך ל: $S \to C_bE,\ E \to SC_b$. לכן יש בדקדוק שלנו $V = \{S, C_a, C_b, D, E\}$ כלומר יש לנו $|V| = 5$ משתנים ולכן החסם המובטח מלמת הניפוח הוא $n \le 2^5 = 32$.

תרגיל: נתונה גרסת למת הניפוח לשפות ח"ה (הכל אותו הדבר מלבד $2 \le |vx|$).
הוכיחו כי השפה $L = \{a^*\} \cup \{b^j a^{n^2} | 0 < j, 1 \le n\}$ אינה ח"ה בעזרת הלמה הזאת.
פתרון: נניח בשלילה כי L ח"ה ויהי n הקבוע המובטח מהלמה. תהי $z = ba^{n^2} \in L$ המקיימת $|z| = n^2 + 1 \ge n$ ולכן קיים פירוק $z = uvwxy$ המקיים את שלוש התנאים $|vwx| \le n$, $2 \le |v| + |x|$. עבור $vx = ba$ ננפח עם $i = 2$ ונקבל $baba^{n^2} \notin L$ כי יצאנו מהפורמט. עבור $vx = a^{|vx|}$ ננפח ב$i = 2$ וממשיך כמו שהוכחנו בהרצאה עבור השפה $L = \{a^{j^2} | j \ge 1\}$ שאינה ח"ה. לכל פירוק מצאנו i טבעי עבורו המילה המנופחת לא בשפה, בסתירה ללמת הניפוח, ולכן השפה אינה ח"ה.

תרגיל: נגדיר מודל חדש: $\tilde{M} = (Q, \Sigma, q_0, \delta, F, \Gamma_1, \Gamma_2, \perp_1, \perp_2)$ (א"מ עם 2 מחסניות).
האוטומט מקבל מילה אם המילה נגמרה והגענו למצב מקבל, או שהמילה נגמרה ושתי המחסניות ריקות. $\delta\left(Q \times (\Sigma \cup \varepsilon) \times \Gamma_1 \times \Gamma_2\right) \to 2^{Q \times \Gamma_1^* \times \Gamma_2^*}$.
בנו אוטומט ממודל זה עבור השפה: $L_2 = \left\{ww^Rww^R\ |\ w \in \{a, b\}^*\right\}$.
פתרון: עבור $\forall \sigma, \sigma_1, \sigma_2 \in \{a, b\}$ מתקיים: $\delta(q_0, \sigma, \perp_1, \perp_2) = (q_0, \sigma\perp_1, \perp_2)$
$$\delta(q_0, \sigma_1, \sigma_2, \perp_2) = (q_0, \sigma_1\sigma_2, \perp_2)\ ,\ \delta(q_0, \varepsilon, \sigma, \perp_2) = (q_1, \sigma, \perp_2)$$
$$\delta(q_1, \sigma, \sigma, \perp_2) = (q_1, \varepsilon, \sigma\perp_2)\ ,\ \delta(q_1, \sigma, \sigma, \sigma_1) = (q_1, \varepsilon, \sigma\sigma_1)$$
$$\delta(q_1, \varepsilon, \perp_1, \sigma) = (q_2, \perp_1, \sigma)\ ,\ \delta(q_2, \sigma, \perp_1, \sigma) = (q_2, \sigma, \varepsilon)$$
$$\delta(q_2, \sigma, \sigma_1, \sigma) = (q_2, \sigma\sigma_1, \varepsilon)\ ,\ \delta(q_2, \sigma, \perp_2) = (q_3, \sigma, \perp_2)$$
$$\delta(q_3, \sigma, \sigma, \perp_2) = (q_3, \varepsilon, \perp_2)\ ,\ \delta(q_3, \varepsilon, \perp_1, \perp_2) = (q_4, \varepsilon, \varepsilon)$$
הסבר: נבנה א"מ המקבל ע"י ריקון 2 המחסניות לשפה. כללי: בעת קריאת w נמלא את המחסנית הראשונה, בעת קריאת w^R נרוקן את המחסנית הראשונה וגם נמלא את השנייה (שימו לב כי, בעת קריאת w נרוקן את המחסנית ה-2 ונמלא את ה-1 ובעת קריאת w^R נרוקן את המחסנית ה-1. נצטרך להשאיר את התחתיות ב-2 המחסניות כדי שנוכל לבצע צעדים נוספים. בסוף נרוקן את שתיהן במסע אפסילון למצב החדש.

דור עזריה - 2022

שפות לא רגולריות שהוכחנו בהרצאות		
$\{a^n b^n \mid n \ge 1\}$	$\{x \in \{a,b\}^* \mid \#_a(x) = \#_b(x)\}$	
$\{xx \mid x \in \{a,b\}^*\}$	$\{a^{k^2} \mid k \in N\}$ ולא ח"ה	$\{a^i b^j \mid i \ne j\}$
$\{a^i b^j \mid i \le j\}$	$\{a^i b^j \mid i \ne j\}$	$\{a^i b^j c^k \mid min(i,j) \le k\}$
$\{a^n c^n\} \cup \{a^n b^n c^n\}$	$\{a^n b^m c^m\} \cup \{a^n b^n c^m\}$	$\{a^n b^{3n} c^n\} \cup \{a^n b^{3n} c^{3n}\}$
$\{a^j b^j c^j \mid j \ge 1\}$ לא ח"ה	$\{ww \mid w \in \Sigma^*\}$ לא ח"ה	

הוכחות מהרצאות

טענה: אם A אס"ד המקבל את L אזי $index(R_L) \le |Q_A|$, שהוא מספר מצבי האוטומט A_L מהוכחת משפט נרוד. כלומר, אם יש לנו אוטומט מהשפה L, אז כמות המצבים שלו מינימלית. **הוכחה**: לפי ההגדרה $L = L(A)$ ולכן R_A מעדן את L.
ממשפט האפיון של R_L מתקיים כי R_A מעדן את R_L, ולכן $index(R_L) \le index(R_A)$.
ובקצרה $|Q_{A_L}| = index(R_L) \le index(R_A) \le |Q_A|$ ובקצרה $|Q_{A_L}| \le |Q_A|$, כנדרש.

טענה 2 של Hoptoft: אם $\tilde{q} \in \tilde{F}$ אזי $\tilde{q} \subseteq F$. **הוכחה**: מההגדרה של $\tilde{F}$ נובע כי אם $\tilde{q} \in \tilde{F}$ אזי $\exists q \in \tilde{q}$ כך ש-$q \in F$. נניח בשלילה כי קיים $p \in \tilde{q}$ אך $p \notin F$. מכאן p, q ניתנים להפרדה ע"י ϵ, כלומר, לפי טענה 1, הם צריכים להשתייך למחלקות שקילות שונות של $\bar{R}$ (ולהיכנס ל-R כבר באיטרציה 0). בסתירה לכך ש-$p, q \in \tilde{q}$, באותה מחלקת שקילות של $\bar{R}$.

טענה 3 של Hopcroft: $\tilde{\delta}$ מוגדרת היטב (מוגדרת באותה צורה לכל מועמד $q \in \tilde{q}$ שאנו בוחרים). **הוכחה**: הגדרנו $\tilde{\delta}(\tilde{q}, \sigma) = [\delta(q,\sigma)]_{\bar{R}}$ כך ש- $q \in \tilde{q}$. נראה שאם $p, q \in \tilde{q}$, אזי $[\delta(q,\sigma)]_{\bar{R}} = [\delta(p,\sigma)]_{\bar{R}}$. נניח בשלילה כי זה לא נכון, נסמן $\delta(q,\sigma) = q', \delta(p,\sigma) = p'$ ונניח כי $[p']_{\bar{R}} \ne [q']_{\bar{R}}$. ההנחה הזו שקולה לכך ש-$(p', q') \notin \bar{R} \leftrightarrow (p', q') \in R$. לכן, מטענה 1 קיים z המפריד בין p', q'. אבל אז σz מפריד בין p ל-q בסתירה לכך ש-$(p, q) \in \bar{R}$.

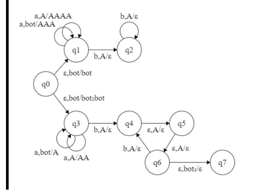
סגירות ח"ה: $L_1 \cup L_2$ – בהינתן L_1, L_2 שפות ח"ה, קיימים להן דקדוקים חסרי הקשר $G_1 = (V_1, T_1, S_1, P_1),\ G_2 = (V_2, T_2, S_2, P_2)$ אזי הדקדוק הבא מתאים לשפה $L_1 \cup L_2$: $G_{1\cup2} = (V_1 \cup V_2 \cup \{S\}, T_1 \cup T_2, S, P_1 \cup P_2 \cup (S \to S_1 | S_2\}$
הוכחה: כיוון ראשון: תהיי $w \in L_1 \cup L_2$. יש להראות כי $w \in L(G_{1\cup2})$ $w \in L_1 \to S_1 \Rightarrow^*$ או $w \in L_2 \to S_2 \Rightarrow^*$ היות וכללי הגזירה של $G_{1\cup2}$ מכילים את כל כללי הגזירה מ - G_2, G_1 $S \to S_1 | S_2$ נקבל כי $w \in L_1 \cup L_2$. כיוון שני: תהיי $w \in L(G_{1\cup2})$. היות וכל המילים הנגזרות מ- $G_{1\cup2}$ הן מהצורה - $S \to S_2 \Rightarrow^* w$ אזי $w \in L_1$ או $w \in L_2$.

הוכחת שקילות $L_\varepsilon(M) \equiv L_f(M)$ **כיוון א**: בהינתן $L_f(M)$ נראה שניתן לבנות M' מ $L_\epsilon(M)$ תהיי M_1 עבורה $L = L_f(M)$ ו- $M_1 = (Q, \Sigma, \Gamma, q_0, \delta, F, \perp)$ אזי קיים א"מ M_2 כך ש- $L = L_\epsilon(M_2)$. רעיון הבנייה : הא"מ M_2 יתנהג כמו M_1, אך כאשר תיכנס למצב מקבל, נכניס את M_2 למצב שמרוקן את המחסנית. נצטרך לזהות מתי M_1 רוקנה את המחסנית. ולכן נשתמש ב"תחתית כפולה". לכן בתחילת הריצה M_2 תכיל את התחתית $\perp_2$, ובצעד הראשון היא תכניס את $\perp_1$. לכן יהיה גם q_0 שביצע את ההכנסה הזו, לפני הרצת M_1.
לכן נקבל את א"מ הבא :
$M_2 = Q \cup \{q_e, q_0'\}, \Sigma, \Gamma \cup \{\perp_2\}, \delta', q0', \perp_2, \phi)$ כאשר $q_0' \notin Q$. q_e וגם $\perp_2 \notin \Gamma$ נגדיר את δ': $\delta'(q_0', \epsilon, \perp_2) = \{(q_0, \perp_1, \perp_2)\}$. אם $q \notin F$ או $\sigma \notin \epsilon$ אז M_2 מתנהג כמו M_1, כלומר לכל $Z \in \Gamma$ $\delta'(q, \sigma, Z) = \delta(q, \sigma, Z)$. אם $q \in F$ אז$\forall Z \in \Gamma \cup \{\perp_2\}$ נגדיר $\delta'(q, \epsilon, Z) = \delta(q, \epsilon, Z) \cup \{(q_e, \epsilon)\}$ (כלומר, אפשר להתקדם רגיל כמו ב M_1 ואפשר ללכת למצב בו מרוקנים את כל המחסנית). $\forall Z \in \Gamma \cup \{\perp_2\}$ נגדי= $\delta(q_e, \epsilon, Z) = (q_e, \epsilon)$ (נשארים במקום ומרוקנים). נוכיח כי $L_f(M_1) = L_\epsilon(M_2)$. תחילה נוכיח כי :
$L_f(M_1) \subseteq L_\epsilon(M_2)$. תהיי $x \in L_f(M_1)$ כלומר קיים חישוב $(q_0, x, \perp_1 \vdash^* (p, \epsilon, \gamma)$ עבור $p \in F$ ו- $\gamma \in \Gamma^*$. נמשיך את המסלול של x עד שיתקבל גם M_2 כך : עפ"י כלל א' (של δ) :
$(q_0', \epsilon, \perp_2) \vdash_{M_2} (q_0, x, \perp_1, \perp_2)$ עפ"י כלל ב', M_2 מתנהג כמו M_1 בכל קריאת המילה x,
ולכן $(q_0, x, \perp_1) \vdash^*_{M_2} (p, \gamma, \perp_2)$. לכן משם נעבור למצב q_e בעזרת כלל ג' (ריקון אות
ונמשיך לרוקן עפ"י כלל ד': $(p, \epsilon, \gamma, \perp_2) \vdash^*_{M_2} (q_e, \epsilon, \epsilon)$. לסיכום מתקיים
$(q'_0, x, \perp_2) \vdash^*_{M_2} (q_e, \epsilon, \epsilon)$ תהיי $x \in L_\epsilon(M_2)$ לכן בקריאת x המחסנית התרוקנה לחלוטין.
הדרך היחידה לעשות זאת היא ע"י הגעה למצב q_e (שכן שאר צעדי M_1 לא יכולים לרוקן את $\perp_2$). לכן סדרת הצעדים ש- M_2 מבצע על x , פרט לצעד הראשון , קיימת גם ב- M_1 ולכן $x \in L_f(M_1)$. לסיכום : $L_\epsilon(M_2) = L_f(M_1)$.

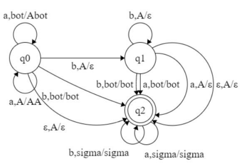
שקילות ב"ר: יהיו $r_1 = (0 * 1)*,\ r_2 = \varepsilon + (0 + 1) * 1$. נוכיח כי ביטויים אלה מייצגים אותה שפה. **כיוון 1:** נניח כי $w \in L[r_1]$. אזי $w = \varepsilon$ או $\exists w_1,..., w_n \in L[0 * 1]$ כך ש: $w = w_1 \cdot .. \cdot w_n$. אם $w = \varepsilon$ אז $w \in \{\varepsilon\} \cup \{0,1\} * \{1\} = L[r_2]$. אחרת, w מסתיימת ב-1. ולכן $w \in \{\{0,1\} * 1\} \subseteq L[r_2]$. **כיוון 2:** נניח כי $w \in L[r_2]$. אם $w = \varepsilon$ אז בוודאי $w \in (\{0\}\{1\})* = L[r_1]$. אם $w \ne \varepsilon$ נוכל לכתוב $w = x1$ כאשר $x \in L[(0 + 1)*]$. נניח כי ב-x יש k מופעים של 1. במקרה זה נוכל לכתוב:
$x = y_1 1 y_2 1 .. y_k 1 y_{k+1}$. כאשר לכל $1 \le n \le k + 1$ מתקיים $y_i \in \{0\}*$. אולם אז $w = (y_1 1)(y_2 1)...(y_k 1)(y_{k+1} 1)$ ולכן $w \in L[r_1]$.

תרגיל: הי דקדוק ח"ה G בו כל הכללים מהצורה $A \to \sigma$, $A \to BCD$, ו-$\epsilon \notin L(G)$, ותהי מילה w. **א**. כתבו אלגוריתם המכריע האם $w \in L$ מבלי לעבור לצורה הנורמלית של חומסקי. ציינו במפורש את אורך עץ הגזירה המבוקש. **פתרון**: בשאלה המקורית נעזרים בצורה הנורמאלית של חומסקי, ואז אורך עץ הגזירה הוא $2|w| - 1$. כאן כללי הגזירה מוגדרים אחרת, ולכן אורך הגזירה אחר. בכל צעד גזירה שבו מקבלים משתנים, נקבל 2 משתנים חדשים. לכן נצטרך סדרה של $\frac{|w|-1}{2}$ צעדי גזירה בה מקבלים משתנים, ואז עוד $|w|$ צעדי גזירה בהם מחליפים משתנה באות. לכן, אורך עץ הגזירה הנדרש הוא $\frac{3|w|-1}{2}$. להלן אלגוריתם: עבור על כל סדרות עצי הגזירה באורך $\frac{3|w|-1}{2}$ – אם אחת מהן גוזרת את w החזר $True$. אחרת החזר $False$.

תרגיל: בנו א"מ לשפה:
$L = \{a^n b^{3n} \cup a^{3n} b^n \mid 1 \le n\}$. **פתרון**:



תרגיל: בנו א"מ לשפה:
$L = \{a^n b^m \mid n \ne m\}$. **פתרון**:



תרגיל: יהי G דח"ה עבורו S אינו משתנה אפיס. כתבו אלגוריתם המכריע האם $50 \le |L(G)|$ אין צורך להוכיח את נכונותו, אך הסבר קל יועיל... **פתרון**: בשאלה מבקשים לבדוק עבור דקדוק כללי כמה מילים הוא יכול לייצר (לפחות). הרעיון הוא לעבור על כל קומבינציות הגזירה. כל קומבינציית גזירה מייצרת עץ גזירה אחד שיכול להסתיים במילה טרמינאלית או בתבנית שיש בה עדיין משתנים. כדי שכל גזירה "תנפח" את המילה אנו נמיר תחילה את הדקדוק לצורה הנורמאלית של חומסקי: **1.** נבטל משתנים לא ישיגים **2.** נבטל משתנים לא טרמינאליים **3.** נבטל כללי אפסילון **4.** נבטל כללי יחידה **5.** נהפוך את הכללים לצורה של חומסקי: $A \to XY$ או $A \to a$ (או 2 משתנים או תו בודד). ככה כל גזירה מעלה את אורך המילה בלפחות אחד. כעת, עוברים על כל עצי הגזירה האפשריים בגובה עד $|V| + 1$ (כמות המשתנים + 1) כי במצב כזה ייתכן ולפי שובך היונים נצטרך לחזור על משתנה יותר מפעם אחת ואז זה אומר שנקבל אינסוף מילים (כל המשתנים הם טרמינאליים ולכן חזרה על משתנה פירושה שניתן לחזור עליו שוב ושוב ובכל פעם לקבל בסוף מילה שונה).
אם לא מצאנו מעגל (משתנה שחוזר על עצמו בעץ אחד), אז פשוט נעבור על כל העצים הללו ונמנה כמה מילים שונות נוצרו, אם הגענו ל 50 - נחזיר true ואם לא - נחזיר false.

תרגיל: **(א)** למה: כל מספר $n \ge 6$ יכול להירשם כסכום של מספר ראשוני על מספר פריק. היזרו בטענה לעיל כדי להוכיח שהשפות הלא רגולריות אינן סגורות לשרשור. חובה להיעזר בטענה. **(ב)** הוכיחו/הפריכו: השפות הלא רגולריות סגורות למשלים. **פתרון א'**: הוכחה: ניקח את השפות הלא רגולריות הבאות: $L_1 = \{a^p \mid p\ is\ prime\},\ L_2 = \{a^c \mid c\ is\ composite\}$
ומכאן: $L_1 L_2 = \{a^{p+c} \mid p\ is\ prime\ and\ c\ is\ composite\}$ לפי הטענה כל החזקות החל מ 6 נמצאות בקבוצה. ולכן: $r = a^6 a^*$ הוא ביטוי רגולרי עבור השפה ולכן השפה רגולרית למרות ש 2 השפות הנ"ל אינן רגולריות. **פתרון ב'**: הוכחה: תהי L שפה לא רגולרית ונניח בשלילה ש $\bar{L}$ כן רגולרית. מכאן: $L = \bar{\bar{L}}$ היא רגולרית מסגירות רגולריות למשלים. סתירה לכך ש L לא רגולרית.

תרגיל: יהי G דח"ה ותהי $w \in T^*$. להלן אלגוריתם לבדיקת שייכות המילה לדקדוק. **(1)** המירו את הדקדוק לא"מ כפי שלמדנו בהרצאה 12. **(2)** בצע DFS על עץ מסלולי החישוב של האוטומט. אם מצאת מסלול מקבל - קבל. אחרת דחה. **(א)** הראו כי האלגוריתם עוצר. **(ב)** הראו שכאשר האלגוריתם עוצר, הוא מחשב נכון. **פתרון א'**: במקרה וקיימת גזירה רקורסיבית של משתנים לצד שמאל (ללא קריאת תווים) אז נוכל לקבל מסלול גזירה אינסופי ובעקבותיו גם מסלול חישוב באוטומט המחסנית שהוא אינסופי ולכן האלגוריתם יסרוק אותו ולא יעצור.
פתרון ב': אם האלגוריתם עוצר אז: מקרה 1 (המילה בשפה): כי מצאנו מסלול חישוב מקבל והיות והא"מ שקול לדקדוק לפי נכונות הבנייה, נקבל שהמילה גם שייכת לשפת הדקדוק (כי קיים מסלול גזירה טרמינלי והמחסנית תתרוקן). מקרה 2 (המילה לא בשפה): סיימנו לסרוק הכל ולא מצאנו מסלול חישוב מקבל ולכן במקרה כזה המילה לא שייכת לשפת האוטומט ולכן גם לא לשפת הדקדוק.

תרגיל: הוכיחו בלמת הניפוח $L = \{a^p:\ p\ is\ prime\}$ לא רגולרית. **פתרון**: נוכיח ש L_3 לא רגולרית ע"י למת הניפוח לשפות רגולריות: נניח בשלילה ש L_3 רגולרית ולכן מקיימת את הלמה. יהי $n \in N$ הקבוע המובטח מהלמה. נבחר: $z = a^p \in L$ כאשר $p \ge n$ ראשוני כלשהו (קיים כזה כי יש אינסוף ראשוניים בעולם) ומתקיים: $|z| = p \ge n$. יהי $z = uvw$ פירוק כלשהו המקיים: $|uv| \le n$. $|v| \ge 1$. מכאן: $v = a^k$ כאשר $1 \le k \le n$.
נבחר: $i = p + 1$ ונקבל: $z' = uv^i w = a^{p+pk} = a^{p(1+k)} \notin L_3$ כי $p(1 + k)$ הוא מספר פריק.

תרגיל: הוכיחו שהשפה $L = \{a^n b^m c^k \mid k \ge min(m,n)\}$ לא רגולרית לפי נרוד.
פתרון: נתבונן בקבוצת המילים: a^i לכל $i \in N$. נראה שכל 2 מילים בקבוצה הנ"ל ניתנות להפרדה ביחס R_{L_2} ומזה ינבע שיש ליחס אינסוף מחלקות שקילות ולפי משפט נרוד השפה אינה רגולרית. עבור $x = a^i,\ y = a^j$ כאשר $i > j$. נבחר את הסיפא: $z = b^i c^j$ ונקבל: $xz = a^i b^i c^j \notin L_2$ כי $j < min(i,i) = i$. ו-$yz = a^j b^i c^j \in L_2$ כי $j \ge min(i,j) = j$.
ולכן הוכחנו שכל 2 מילים בקבוצה אכן ניתנות להפרדה ושייכות למחלקות שונות.

תרגיל: יהי דקדוק ליניארי ימני G (כל הכללים מהצורות $A \to \sigma B,\ A \to \sigma, S \to \varepsilon$) ותהי סדרת גזירה $S \to^* \alpha$. הוכיחו כי $\alpha = wB$ או $\alpha = w$ עבור $B \in V,\ w \in T^*$ כלשהם.
פתרון: נוכיח באינדוקציה על מספר צעדי הגזירה. בסיס: $n = 1$. ע"י צעד גזירה אחד ניתן לקבל: $S \to \varepsilon$ (מילה) או $S \to \sigma$ (מילה) או $S \to \sigma X$ (מהצורה wB) צעד: נניח שהטענה נכונה לכל תבנית הנגזרת לאחר n צעדים ונוכיח עבור תבניות הנגזרות לאחר n+1 צעדים.
תהי α תבנית כך ש: $S \Rightarrow^{n+1} \alpha$. מכאן: צעד הגזירה האחרון היה מהצורה: $A \to \sigma B$ או $A \to \sigma$ או $S \to \varepsilon$. לפי הנחת האינדוקציה נקבל ש: $S \Rightarrow^n wA \to w\sigma B = xB$ או $S \Rightarrow^n wA \to w\sigma = x$ או $S \Rightarrow^n wS \to w = x$. ובכל מקרה קיבלנו תבנית כנדרש.

תרגיל: יהי דקדוק G עם אוסף כללי הגזירה הבאים: $S \to ab|a^2b|a^3b|aSb|a^2Sb|a^3Sb$
(א) קבעו מהי השפה $L(G)$. **(ב)** הוכיחו $L(G) \subseteq L$.
פתרון א': $\{a^n b^k : 1 \le k \le n \le 3k\}$.
פתרון ב': נוכיח טענה חזקה יותר: לכל $n \ge 1$ ולכל צורה פסוקית $\alpha \in (V \cup T)^*$ מתקיים $S \to^n \alpha$ אם"ם α מהצורות $a^n S b^k$ או $a^n b^k$ (עבור $k \le n \le 3k$). לאחר שנוכיח את הטענה החזקה נקבל כי $L(G) \subseteq L$ עבור $\alpha = a^n b^k$ שכן הצורה השניה לא טרמינלי. נוכיח את הטענה החזקה באינדוקציה על n: **בסיס**: $n = 1$ נקבל את אחת משש הגזירות הבאות: $S \to aSb,\ a^2Sb,\ a^3Sb,\ ab,\ a^2b,\ a^3b = \alpha$ כנדרש. **הנחה**: נניח כי הטענה נכונה עבור n צעדי גזירה. **צעד**: נוכיח עבור $n + 1$ צעדים. כלומר, קיימת סדרת הגזירה הבאה: $S \to^n \alpha \to \beta$. לפי הנחת האינדוקציה $\alpha = a^n S b^k$ (שכן אחרת קיבלנו מילה טרמינלית ואז לא ניתן להמשיך את הגזירה). לכן, עבור $t = 1, 2, 3$ נקבל: $a^n S b^k \to a^{n+t} S b^{k+1}$ או $a^n S b^k \to a^{n+t} b^{k+1}$. כנדרש.

ב. אם נאפשר גם $S \to \epsilon$ ניקלע לבעיה. מהי, וכיצד ניתן לפתור אותה?
פתרון: הבעיה היא שכעת לא נוכל לשלוט על אורך עצי הגזירה – יתכן והמילה w תיגזר ע"י עץ ארוך מאוד, שבו חלק מהמשתנים הופכים ל ϵ במקום לטרמינל.כיצד נפתור זאת? נוכל לבצע "סילוק כללי אפסילון" מהדקדוק, אך אז הוא כבר לא יישאר בצורה $A \to BCD$, אלא יתכנו 1, 2, 3 משתנים, או טרמינל בצד ימים. כללי היחידה מובילים לאותה בעיה של עומק עץ לא חסום.

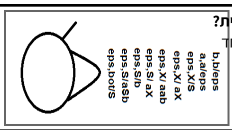
לדוגמה, בדקדוק $S \to SSS|a|\epsilon$ נקבל $S' \to S|\epsilon S \to SSS|SS|S|a$ כדי להתגבר על הבעיה, לאחר סילוק כללי אפסילון נצטרך לסלק גם כללי יחידה. אמנם לא נוכל לתת חסם מדויק על עומק העץ, אבל יהיה לנו חסם עליון של $2|w| - 1$ כמו בחומסקי רגיל.

חלאס.

תרגיל: נגדיר את המודל $4regNFA$ בצורה הבאה: המודל הינו אוטומט מסוג NFA עבור בכל צעד חיובי ניתן ללכת בדיוק לארבעה מצבים שונים. הוכיחו כי מודל זה שקול ל- NFA רגיל. הנחה: הראו שקילות ל-NFA. המודל $4regNFA$ עונה על ההגדרה הבאה: $\forall q \in Q,\ \forall \sigma \in \Sigma: |\delta(q,\sigma)| = 4$. **פתרון**: נראה שקילות בין $4regNFA$ לבין DFA וכבר הוכחנו בהרצאה $NFA \equiv DFA$ ובכך נסיים. כיוון ראשון: בהינתן $4regNFA$ ניתן לבנות DFA שקול בעזרת האלגוריתם של אוטומט חזקה. כיוון שני: בהינתן $A = (Q, \Sigma, q_0, \delta, F)$ נגדיר: $regNFA = (Q', \Sigma, q_0, \delta', F)$ כאשר $Q' = Q \cup \{q_{pit1}, q_{pit2}, q_{pit3}, q_{pit4}\}$ בצורה הבאה: $\forall q \in Q, \forall \sigma \in \Sigma$ מתקיים: $\delta'(q,\sigma) = \{\delta(q,\sigma),\ q_{pit1}, q_{pit2}, q_{pit3}\}$. וגם לכל $\forall q \in \{q_{pit1}, q_{pit2}, q_{pit3}, q_{pit4}\}$ ולכל $\sigma \in \Sigma$ מתקיים: $\delta(q,\sigma) = \{q_{pit1}, q_{pit2}, q_{pit3}, q_{pit4}\}$.

תרגיל מבחן: למת ניפוח לח"ה. השפה: $L = \{a^n b^{3n} c^{9n} \mid 1 \le n\}$. קבעו אם רגולרית או ח"ה או ת"ה. **פתרון**: לא חסרת הקשר. נוכיח שהיא לא מקיימת את למת הניפוח לחסרות הקשר. נניח בשלילה ש L חסרת הקשר ולכן מקיימת את למת הניפוח לחסרות הקשר. יהי $n \in N$. נבחר את המילה: $z = a^n b^{3n} c^{9n} \in L$ וגם $|z| = 13n \ge n$.
לכן קיים פירוק: $z = uvwxy$ המקיים: **(1)** $|vwx| \le n$ **(2)** $|vx| \ge 1$.
נחלק למקרים: **מקרה 1**: אם $vx = a^k$ עבור $1 \le k \le n$ (לפי (1) + (2)). נבחר $i = 0$ ונקבל: $z' = uv^0wx^0y = a^{n-k} b^{3n} c^{9n} \notin L_3$. **מקרה 2**: אם $vx = b^k$ עבור $1 \le k \le n$ (לפי (1) + (2)). נבחר $i = 0$ ונקבל: $uv^0wx^0y = a^n b^{3n-k} c^{9n} \notin L_3$.
מקרה 3: אם $vx = c^k$ עבור $1 \le k \le n$ (לפי (1) + (2)). נבחר $i = 0$ ונקבל: $z' = uv^0wx^0y = a^n b^{3n} c^{9n-k} \notin L_3$. **מקרה 4**: אם $vx = a^k b^m$ עבור $1 \le k, m \le n$ (לפי (1) + (2)). נבחר $i = 0$ ונקבל: $uv^0wx^0y = a^{n-k} b^{3n-m} c^{9n} \notin L_3$. **מקרה 5**: אם $vx = b^k c^m$ עבור $1 \le k, m \le n$ (לפי (1) + (2)). נבחר $i = 0$ ונקבל: $z' = uv^0wx^0y = a^n b^{3n-k} c^{9n-m} \notin L_3$. בכל מקרה קיבלנו סתירה ולכן השפה לא מקיימת את הלמה ולכן לא ח"ה.

איך ממירים דקדוק לאוטומט מחסנית?
הרעיון הוא לבנות אוטומט עם מצב אחד המקבל ע"י ריקון כאשר **כללי הגזירה** **יהיו** המעברים: $S \to aSb \mid b \mid aX$
$X \to aab \mid aX \mid S$



תרגיל: תהא שפה L ומחלקת שקילות S של R_L. הוכיחו: $\min_{w \in S} |w| < index(R_L)$.
פתרון:נחלק למקרים: **(1)** אם L לא רגולרית אז לפי משפט נרוד נובע ש: $index(R_L) = \infty$ וגודל המילה המינימלית ב-S הוא מספר כלשהו שבטוח קטן מאינסוף. **(2)** אם L רגולרית אז קיים לה אס"ד (DFA) מינימלי עם בדיוק $index(R_L)$ מצבים (הוכח בהרצאה). תהי $m \in S$ המילה המינימלית באורכה ב-S. נניח בשלילה ש $|m| \ge index(R_L)$. מכיוון שאורך המילה גדול או שווה לכמות המצבים באס"ד הנ"ל אז לפי עקרון שובך היונים, כשנקרא את המילה m מ-q_0 בהכרח נחזור לפחות פעמיים למצב שכבר היינו בו (יש מעגל). נסמן: $m = uvw$ כאשר u היא הקריאה עד המעגל, v הוא המעגל, w אחרי המעגל. מכיוון שיש מעגל אז $|v| \ge 1$ אבל מכאן, המילה uw תגיע בדיוק לאותו מצב סופי כמו m. לכן, לפי הנתון שהאוטומט הוא מינימלי נקבל ש m, uw נמצאות באותה מחלקת שקילות כי באס"ד מינימאלי כל 2 מילים שמגיעות לאותו מצב נמצאות באותה מחלקת שקילות. אבל uw היא מילה קצרה יותר מ-m בסתירה לכך ש-m הייתה המינימאלית באורכה באותה המחלקה.

תרגיל: לכל מספר טבעי $n \in Z^+$ קיימים $a, b, c, d \in Z^+ \cup 0$ כך שמתקיים: $n = a^2 + b^2 + c^2 + d^2$. הוכיחו בעזרת משפט זה כי השפות הלא רגולריות לא סגורות לשרשור. **פתרון**: דוגמא נגדית: $L = \{1^{n^2} \mid n \in N\}$ - לא רגולרית.
מכאן: $L \cdot L = \{1^{a^2+b^2} \mid a, b \in N\}$ - לא רגולרית (כי אחרת כבר הפרכנו את הטענה)
ולכן: $L^4 = L^2 \cdot L^2 = \{1^{a^2+b^2+c^2+d^2} \mid a,b,c,d \in N\} = \{1^n \mid n \in N\}$ שפה רגולרית.

תרגיל: יהיו A_1, A_2 שני DFA לשפות L_1, L_2. כתבו אלגוריתם המכריע האם קיימת מילה w השייכת לשפה $L_2 \backslash L_1$. **פתרון**: האלגוריתם: נבנה אוטומט לשפה המשלימה של L_1 (קיים כזה כי שפות רגולריות סגורות למשלים) נבנה אוטומט מכפלה ל $L_2 \cap \bar{L_1}$ (השקול להפרש) נסרוק את האוטומט החל מהמצב ההתחלתי ונבדוק האם קיים מסלול שמוביל אותנו למצב מקבל.

תרגיל: תהי $Shuffle(L_1, L_2) = \{u_1 v_1 .... u_n v_n \mid \forall i \in [1,n]: u_i \in L_1, v_i \in L_2\}$.
הוכיחו/הפריכו: **(א)** אם L_1, L_2 רגולריות אזי גם $Shuffle(L_1, L_2)$ רגולריות. **(ב)** אם $Shuffle(L_1, L_2)$ רגולרית אז גם L_1, L_2 רגולריות. **פתרון**: **(א)** נשים לב שמתקיים: $shuffle(L_1, L_2) = (L_1 \cdot L_2)^*$ ולכן מסגירות לשרשור ולאיטרציה נקבל ש $shuffle(L_1, L_2)$ רגולרית. **(ב)** לא נכון. דוגמא נגדית:
$L_1 = \{w \in \{a,b\}^* \mid \#_a(w) \ne \#_b(w)\},\ L_2 = \{w \in \{a,b\}^* \mid \#_a(w) = \#_b(w)\}$.
2 השפות לא רגולרית. מכאן: $a \in L_1 \cdot L_2$ כי $a \in L_1, \varepsilon \in L_2$. וגם $b \in L_1 \cdot L_2$ כי $b \in L_1, \varepsilon \in L_2$. ולכן: $shuffle(L_1, L_2) = (L_1 \cdot L_2)^* = \{a,b\}^*$ (שפת כל המילים מעל a,b) שהיא רגולרית.

תרגיל: $L_1 = \{a^n b^m \mid n \ne m\}$. קבעו אם רגולרית, ח"ה או ת"ה. **פתרון**:
השפה L_1 היא כן חסרת הקשר ולא רגולרית. הוכחת אי רגולריות: נשתמש במשפט נרוד.
נראה שיש ליחס R_{L_1} אינסוף מחלקות שקילות ואז ממשפט נרוד נסיק ש L_1 אינה רגולרית.
נתבונן בקבוצת המילים הבאה: a^n לכל $n \in N$. נראה שכל 2 מילים בקבוצה ניתנות להפרדה זו מזו. עבור $x = a^i,\ j = a^j$ כאשר $i \ne j$ נבחר את הסייפא: $z = b^i$ ונקבל: $xz = a^i b^i \notin L_1$ אבל $yz = a^j b^i \in L$. ולכן השפה אינה רגולרית. הוכחה ש L_1 כן חסרת הקשר: נראה דקדוק עבור השפה: $G = (V = \{S, A, B\}, \Sigma = \{a,b\}, S, P)$
$S \to aSb \mid aA \mid Bb, \quad A \to aA \mid \varepsilon \quad , \quad B \to Bb \mid \varepsilon$

תרגיל: בנו ב"ר לשפה $L = \{a^n b^m \mid n + m \equiv 3 (mod\ 5)\}$.
פתרון: $L = (a^5)^* (a^3 + a^2b + ab^2 + b^3 + a^4a^4)(b^5)^*$.