אוטומטים ושפות פורמליות

עומר גולד

תקציר

חוברת זו מבוססת על תרגולים שהעברתי בשנת תשע"ב (2012) באוניברסיטת בר־אילן.

אני מקווה שחוברת זו תוכל לעזור לסטודנטים, ולכל מי שמתעניין בתאוריה של מדעי המחשב.
החוברת מכילה הגדרות ומגוון של דוגמאות ותרגילים על הנושאים השונים שנלמדו בקורס. מכיוון שהחוברת מבוססת על תירגולים, מספר הוכחות של משפטים מרכזיים אינן מופיעות בפרקי השיעורים, אך על־מנת לא להחסיר אותן מהקוראים, הוכחות אלו מופיעות כפי שנלמדו בהרצאות של פרופ' משה קופל בפרק 12.

המקורות העיקריים מהם לקוח החומר שמוצג כאן ושהועבר על־ידי בתרגולים הם:

1. Introduction to Theory of Computation (Michael Sipser)
2. Computability, Complexity, and Languages (Martin D. Davis, Ron Sigal, Elaine J. Weyuker)
3. אוטומטים ושפות פורמליות (האוניברסיטה הפתוחה)

תודה מיוחדת לאליה גריידי על סיכום החומר ב־ LaTeX במהלך הסמסטר.

מבוא

מה זה אוטומט? מהי שפה פורמלית? לאיזה תחומים זה קשור? למה (ולמי) זה טוב?

קורס זה שייך לענף "חישוביות". תורת החישוביות היא הבסיס למדעי המחשב, והיא עוסקת במודלים לחישוב ובפונקציות הניתנות לחישוב במסגרתם. השאלה הבסיסית בתורת החישוביות היא: מה מחשבים יכולים לחשב, ומה לא?

כולנו יודעים מהו מחשב ־ אותה מכונה היושבת על שולחן המשרד בעבודה, או בבית ואפילו בכל טלפון סלולרי. מכונות אלה הינן מורכבות מאד, ואין לנו יכולת לנתח אותן באופן מתמטי מדוייק ולשאול מה הן מסוגלות לעשות ומה לא.

כדי להיות מסוגלים לדון בשאלות כנ"ל דרוש לנו מודל שהוא מצד אחד פשוט, ומצד שני מסוגל לבצע כל פעולה שמחשב "רגיל" יכול לבצע. נוסף על כך, על־מנת שנוכל להגדיר ולהוכיח באופן מדוייק את היכולת של המחשב, המודל שלנו חייב להיות מתמטי.

בקורס הזה נגדיר מודלים מתמטיים של מכונות חישוב למיניהן. נתחיל עם מכונות מאד פשוטות, על־מנת לבסס את העקרונות ולהרחיבן לאט לאט עד שנוכל להגדיר מודל מתמטי שעדיין יהיה פשוט, ובנוסף יהיה שקול (מבחינת "כח החישוב" שלו) לכל מחשב "רגיל", מודל זה נקרא "מכונת טיורינג" ועליו תדברו בהרחבה גם בקורס "חישוביות".

כמו כן, נסביר מהי שפה פורמלית ונראה בקורס את הקשר בין מודל חישובי לשפה פורמלית. מסתבר שניתן לאפיין משפחות של שפות לפי מודל חישובי "המכריע" אותן, כמו כן נראה כי ייתכנו מס' מודלים שונים שקולים אשר "מכריעים" את אותה משפחה של שפות.

בהקשר זה ניתן לומר שקיימת הירככיה המציינת את "כח החישוב" הנדרש על־מנת "להכריע" שפה ממשפחה מסוימת. אפיון כזה הוא חשוב למחקר תיאורתי במדעי המחשב, למשל נשאלת השאלה מהי "הסיבוכיות" הנדרשת על־מנת להכריע שפה ממשפחה מסוימת? האם יש שפה שלא ניתן להכריע?

על־מנת לענות על שאלות אלו, נצטרך קודם כל בסיס מתמטי, אח"כ נתחיל ללמוד על משפחות של שפות ומודלים "חלשים" ועם הזמן נעלה בהירככיה ונגיע למודלים "חזקים" יותר, כשבסוף נגיע ל"מכונת טיורינג".

כמו כן, ישנם גם שימושים מעשיים לכל אחד מהמודלים שנלמד בקורס, למשל: משתמשים באוטומטים סופיים ב"עיבוד טקסטים", "תיאום דגמים" ו"תכנון חומרה". לדקדוקים חסרי־הקשר שימושים מעשיים ניתן למצוא במהדרים ("קומפיילרים"), בתכנון שפות תכנות, ואפילו בבינה מלאכותית, ואילו שפות תלויות־הקשר מעלות שאלות על "סיבוכיות מקום" ובעלות שימוש ב־"עיבוד שפות טבעיות". על שימושים אלו ניתן ללמוד בהרחבה בקורסים היעודיים.

בעיות הכרעה

כדי למדוד "כח חישוב", נתמקד בשאלות הכרעה: שאלות שהתשובה עליהן היא כן/לא.

עבור שאלות הכרעה, ייתכנו קלטים שהתשובה עליהם היא "כן" וקלטים אשר התשובה עליהם היא "לא". אם יש לנו מכונה שעונה נכון על השאלה, אנו אומרים שהמכונה מכריעה את השאלה. קלט עליו המכונה ענתה "כן" נקרא קלט שהתקבל על־ידי המכונה, וקלט שעליו המכונה ענתה "לא" נקרא קלט שנדחה ע"י המכונה.

כדוגמא פשוטה מאד, נניח שהשאלה הינה "האם הקלט הוא שם של חיה?" אזי קלטים שייתקבלו על־ידי המכונה יהיו "כלב", "חתול", וכו'. קלטים שיידחו הם, לדוגמא, "שמים", "אילן" וכד'. ניתן לדבר על קבוצת כל הקלטים המתקבלים, קבוצה זו נקראת גם "שפה" וכל קלט בה נקרא "מילה" בשפה.

למשל, השפה שמתקבלת על־ידי מכונה שמכריעה את השאלה לעיל היא {כלב, חתול, עכבר, ...} או באופן כללי השפה L המוגדרת על־ידי: $L = \{w \mid$ w שם של חיה$\}$.

כמו כן, האם לכל שפה קיימת מכונת טיורינג שמסגולת להכריע אותה? ניתן להראות שלא, אך לא נראה זאת כאן. נושא זה נלמד בהרחבה בקורס "חישוביות".

בדוגמא לעיל, המילים בשפה L הן בעברית, אך כשנלמד על שפות פורמליות, בד"כ נדבר על שפות שלמילים בהן לא בהכרח יש משמעות סמנטית. בפרק הראשון נגדיר באופן מדוייק מהי מילה, ומהי שפה פורמלית.

תוכן עניינים

[1] נקרא גם אוטומט חסום לינארית ־ LBA

חלק I

שיעור 1

1 הגדרות:

א"ב: אוסף (קבוצה) סופי ולא ריק של סימנים/אותיות/תווים. נסמן אותו באות $\sum$.

דוגמאות: $\{0,1,\ldots,9\}$, $\{$א, ..., ת$\}$

מחרוזת/מילה: מילה הינה סדרה סופית של סימנים מא"ב כלשהו. למשל מעל $\{$א,...,ת$\}$. נסמן אותה ב־w.

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

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

שרשור מילים: יהיו w_1,w_2 מילים מעל א"ב $\sum$. אזי $w=w_1w_2$ נקרא שרשור של w_1,w_2
למשל: $w_1=ab$, $w_2=b$ אזי $w=abb$.

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

$\sum^*$:

קבוצת כל המילים מעל א"ב $\sum$. (נשים לב: $\sum^*=\bigcup\limits_{n=0}^{\infty}\sum^n=\sum^0\cup\sum^1\cup\sum^2\cup\ldots$)

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

אות: תו ששייך לא"ב ונסמנוב: $\sigma\in\sum$. מס' המופעים של σ ב־w הוא:$\#\sigma\left(w\right)$

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

2 פעולות על שפות

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

בנוסף נגדיר שרשור שפות כך: יהיו $L_1\subseteq\sum_1^*$ ו־ $L_2\subseteq\sum_2^*$, שרשור $L_1\cdot L_2\subseteq\left(\sum_1\cup\sum_2\right)^*$ מוגדר באופן הבא: $L_1\cdot L_2=\{w=w_1\cdot w_2\,|\,w_1\in L_1,w_2\in L_2\}$

דוגמאות:

1) $\sum_1=\{0,1\}$ $\sum_2=\{0,a,b\}$ $L_1=\{0,1\}$ $L_2=\{0,a,b\}$, $L_1\cdot L_2=\{00,0a,0b,10,1a,1b\}$

2) לכל שפה L: $L\cdot\emptyset=\emptyset\cdot L=\emptyset$

3) לכל שפה L: $L\cdot\varepsilon=\varepsilon\cdot L=L$

הערה: נשים לב ששרשור איננה פעולה קומטטיבית!

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

$$\begin{cases}L^0=\{\varepsilon\}\\ L^1=\{L\}\\ L^n=L^{n-1}\cdot L\end{cases}$$

כלומר, $L^n=\{w=w_1\cdot w_2\cdot\ldots\cdot w_n\,|\,\forall 1\le i\le n\;,w_i\in L\}$

נגדיר: $*$−קליני: $L^*=\bigcup\limits_{n=0}^{\infty}L^n$

דוגמאות: 1) $L=\{0\}$ $L^*=\left\{0^i\,|\,i\ge 0\right\}$

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

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

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

נגדיר L^+ **:** $L^+=\bigcup\limits_{n=1}^{\infty}L^n$

מתי $L^*=L^+$? אם $\varepsilon\in L$ אז $L^+=L^*$.

תרגיל

יהיו L_1,L_2,L_3 שפות מעל $\sum$ הוכח או הפרך:

1) $L_1\cap L_2L_3=(L_1\cap L_2)\,(L_1\cap L_3)$

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

$$L_1\cap L_2L_3=\{ab\}$$

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

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

הוכחה: תהי $w\in L_1(L_2\cup L_3)$ $\Longleftrightarrow$ קיימים $y,z\in\sum^*$: $w=yz$ ומתקיים:

$((z\in L_3\wedge y\in L_1)\vee(z\in L_2\wedge y\in L_1))\Longleftrightarrow((z\in L_3\vee z\in L_2)\wedge y\in L_1)$

$\Longleftrightarrow(w\in L_1L_2\cup L_1L_3)\Longleftrightarrow(w\in L_1L_3\vee w\in L_1L_2)$ כנדרש.

3) $(L_1\cup L_2)L_3=L_1L_3\cap L_2L_3$

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

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

$L_1L_3\cap L_2L_3=\{abba,aba,cba,ca\}\cap\{aba,aa,cba,ca\}=\{aba,cba,ca\}$

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

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

$L_1L_3\cap L_2L_3=\sum^*\cap L_2L_3\neq\emptyset$

הסבר: מכיוון ש־L_2,L_3 אינן קבוצות ריקות, שרשור L_2L_3 אינו קבוצה ריקה ולכן $\sum^*\cap L_2L_3\neq\emptyset$

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

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

תרגיל:

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

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

אם $y=\varepsilon$ אזי כמו מקודם.

אחרת $x,y\neq\varepsilon$ ונסמן $x=x_1x_2\ldots x_n$, $y=y_1y_2\ldots y_m$ ומתקיים:

$(xy)^R=(x_1x_2\ldots x_2y_1y_2\ldots y_m)^R=y_my_{m-1}\ldots y_1x_nx_{n-1}\ldots x_1=y^Rx^R$ כנדרש.

תרגיל:

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

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

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

פתרון: נגדיר $L_k = \{0^n \,|\, n \geq k\}$ ו־ $A = \bigcup\limits_{k=0}^{\infty} \{L_k\} = \{L_0, L_1, L_2, \ldots\}$

הוכחת תנאי 1: נתסכל בתת־קבוצה סופית של A , $A' \subseteq A$ $A' = \{L_{i_1}, L_{i_2}, \ldots, L_{i_m}\}$ נניח בה"כ ש־i_m הוא האינדקס הגדול ביותר. אזי חיתוך כל איברי A' הוא $L_{i_m} = \{0^n \,|\, n \geq i_m\}$ ואכן אינו ריק, כלומר $\bigcap\limits_{L \in A'} L \neq \emptyset$

הוכחת תנאי 2: נסתכל בתת־קבוצה אינסופית כלשהי $A^{''} \subseteq A$. נניח בשלילה ש־$\bigcap\limits_{L \in A^{''}} L \neq \emptyset$ אזי קיימת מילה $w = 0^i$ $(i \geq 0)$ ב־ $\bigcap\limits_{L \in A^{''}}$ מכיוון ש־$A^{''}$ אינסופית אזי קיים $j > i$ כך ש־ $L_j \in A^{''}$

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

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

חלק II

שיעור 2

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

תרגיל:

הוכח: $(L^*)^* = L^*$

כיוון א': $(L^*)^* \subseteq L^*$

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

$\Leftarrow$נשים לב כי $x = y_1 \ldots y_n$ כך ש־$y_i \in L^*$, $\forall 1 \leq i \leq n$ $\Leftarrow$ לכל y_i קיים m_i: $y_i \in L^{m_i}$ $\Leftarrow$ $y_i = z_{i_1} z_{i_2} \ldots z_{i_{m_i}}$ כך ש־$z_{i_j} \in L$

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

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

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

3 אוטומט סופי דטרמיניסטי

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

דוגמא:

$$\sum = \{a,b\}$$
$$Q = \{q_0, q_1\}$$
$$q_0$$
$$F = \{q_0\}$$

את δ אפשר לייצג בטבלה באופן הבא:

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

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

הרחבה של δ למילים: נגדיר $\delta^* : Q \times \sum^* \longrightarrow Q$ באופן רקורסיבי:

$$\delta^*(q,\varepsilon) = q$$
$$\delta^*(q,w\sigma) = \delta(\delta^*(q,w),\sigma)$$

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

תרגילים:

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

1) $L_1 = \{w \mid |w| \, mod 2 = 1\}$ (נהפוך את המצב המקבל בדוגמה הקודמת להיות $F = \{q_1\}$

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

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

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

נשים לב כי $L_4 = \overline{L_3} = \sum^* - L_3$.

נבנה את האס"ד כך: $Q_{L_4} = Q_{L_3}, F_{L_4} = Q_{L_3} - F_{L_3}$ ולכל $q \in Q, \sigma \in \sum$: $\delta_{L_3}(q,\sigma) = \delta_{L_4}(q,\sigma)$

5) $L_5 = \left\{wa\sigma \mid w \in \sum^*, \sigma \in \sum\right\}$

6) $L_6 = \left\{\sigma_1 w \sigma_2 \mid w \in \sum^*, \sigma_1, \sigma_2 \in \sum, \sigma_1 \neq \sigma_2\right\}$

7) $L_7 = \{w \mid |w| \, mod 3 = 1\}$

8) $$L_8 = \left\{ w \mid \underbrace{\#a(w)}_{\text{number of appearances}} mod2 = 1 \wedge \#b(w)\, mod2 = 0 \right\}$$

9) $$\sum = \{a,b,c\} \quad , \quad \begin{matrix} L_9 = \left\{a^i b^j c^k \mid i,j,k \geq 0\right\} \\ L_{10} = \left\{a^i b^j c^k \mid i,j,k \geq 1\right\} \end{matrix}$$

10) $L_{10} = \emptyset$. כל אס"ד A כך ש־$F_A = \emptyset$

11) $L = \{\varepsilon\}$ q_0 מקבל, וכל קלט הולך ל־"מלכודת".

12) $L = \sum^*$ .q_0 מקבל, ומחזיר לעצמו.

3.1 שפה רגולרית:

כך שפה L אשר קיים אס"ד A כך ש־$L(A) = L$ נקראת שפה רגולרית. ניתן לומר כי L רגולרית אםם קיים אס"ד A כך ש־$L(A) = L$.

לכן, כל השפות שראינו בתרגול על אס"ד הינן שפות רגולריות.

חלק III

שיעור 3

4 אוטומט מכפלה

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

נראה כיצד לבנות אוטומט A כזה

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

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

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

באופן פורמלי:

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

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

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

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

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

דוגמה: $\sum = \{0,1\}$, $\begin{matrix} L_1 = \{w \mid \#1(w) \, mod2 = 1\} \\ L_2 = \{w \mid \text{w contains substring 000}\} \end{matrix}$

א) בנה אס"דים המקבלים את L_1, L_2

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

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

פתרון: במחברת

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

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

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

טענה: השפות הרגולריות לא סגורות תחת איחוד אינסופי.

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

טענה: השפות הרגולריות לא סגורות תחת חיתוך אינסופי.

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

כעת נגדיר: $L' = \bigcap_{n=0}^{\infty} \overline{L_n} = \sum^* - \{\varepsilon\} - \{a^1 b^1\} - \{a^2 b^2\} - \ldots = \sum^* - \{a^n b^n \mid n \geq 0\} = \overline{\{a^n b^n \mid n \geq 0\}}$

אזי לפי ההנחה שלנו, L' רגולרית, ולכן (שוב מסגירות תחת השלמה) $\overline{L'} = \{a^n b^n \mid n \geq 0\}$ רגולרית, וזו סתירה.

5 אוטומט סופי לא דטרמיניסיטי (אסל"ד)

אסל"ד מוגדר ע"י חמישייה $\{\sum, Q, q_0, F, \delta\}$ כאשר: $F, q_0, Q, \sum$ כמו באס"ד ופונקציית המעברים δ מוגדרת: $\delta : Q \times \sum \longrightarrow \underbrace{2^Q}_{P(Q)}$ כלומר $\delta(q_1, \sigma) = \{q_1, \ldots, q_r\}$ מציין שאם נמצאים במצב q וקוראים את האות σ אז ניתן לעבור לכל אחד מהמצבים $\{q_1, \ldots, q_r\}$.

ההרחבה של δ למילים מוגדרת כך: $\delta^* : Q \times \sum^* \longrightarrow 2^Q$ (לכל $q \in Q$: $\delta^*(q, \varepsilon) = \{q\}$) ולכל $q \in Q$, $w \in \sum^*$, $\sigma \in \sum$ $\delta^*(q, w\sigma) = \bigcup_{q' \in \delta^*(q,w)} \delta^*(q', \sigma)$

זו ההגדרה של אסל"ד ללא מעברי ε.

מעברי ε:

מעבר בו ניתן לעבור ממצב למצב ללא קריאת אות. כלומר, $\delta(q, \varepsilon) = p$ (כאשר $p \neq q$) הינו מעבר ε.

כל אסל"ד שבו מותרים מעברי ε ניתן להמיר לאסל"ד שקול ללא מעברי ε, לכן אנו נשתמש באסל"ד עם מעברי ε. כלומר, $\delta : Q \times (\sum \cup \{\varepsilon\}) \longrightarrow 2^Q$ ובהתאם δ^*.

לכן אצלנו $\delta^*(q, w)$ יהיה אוסף כל המצבים שניתן להגיע אליהם מ־q לאחר קריאת המילה w תוך התחשבות בכל מעבר ε אפשרי.

שפת האוטומט

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

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

מספר הבדלים בין אס"ד לאסל"ד

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

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

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

דוגמא:

תהא $L = \left\{ w1 \,|\, w \in \sum^{*} \right\}$

א) בנה אס"ד המקבל את L.

ב) בנה אסל"ד המקבל את L.

פתרון: במחברת.

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

דוגמאות במחברת

חלק IV

שיעור 4

משפט:

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

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

דוגמה

נתון האסל"ד הבא: ציור במחברת

צייר אס"ד שקול לו.

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

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

6 ביטויים רגולריים

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

1) $\emptyset$ וכל איבר ב־$\sum$ הוא ביטוי רגולרי

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

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

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

סדר פעולות

$*$־ קליני קודמת לשרשור, שרשור קודם לאיחוד. כך נוכל לצמצם את מופעי הסוגריים בביטוי. דוגמה: $((a(b^*)) \cup (b(a^*))) \equiv (ab^* \cup ba^*)$

שרשור נרשום בפשטות כרצף של תווים. דוגמה: $r = (0 \cdot 1 \cdot 0) \equiv 010$

סמנטיקה של ביטויים רגולריים:

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

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

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

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

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

דוגמה

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

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

משפט

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

הוכחה:

$(\Longrightarrow)$ קיים אס"ד A כך ש-$L\left(A\right)=L$

$(\Longleftarrow)$ קיים אסל"ד A' כך ש-$L\left(A'\right)=L$

תרגיל

תהא L שפה רגולרית. נגדיר $L^{R}=\left\{ w^{R}\,|\,w\in L\right\}$ (w^{R} מציין $reverse\left(w\right)$)
הראה שגם השפה L^{R} רגולרית.

הוכחה: יהי r ביטוי רגולרי כך ש-$L\left(r\right)=L$ (קיים כזה, כי L רגולרית). נראה כי לשפה L^{R} יש ביטוי רגולרי r' כך ש-$L\left(r'\right)=L^{R}$ ונסיק מכך שגם L^{R} רגולרית.

הוכיח זאת באינדוקציה <u>שלמה</u> על האורך של r.

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

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

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

מקרה 1: נבחין כי r_{1},r_{2} מרכיבים את r ושניהם באורך 1 לפחות ואילו r מכיל תווים נוספים. מכאן, $\left|r_{1}\right|,\left|r_{2}\right|<n$. לפי הנחה קיימים ביטויים רגולרים r_{1}',r_{2}' עבורם מתקיים $L\left(r_{1}'\right)=L\left(r_{1}\right)^{R},L\left(r_{2}'\right)=L\left(r_{2}\right)^{R}$. נסמן $r'=\left(r_{1}'\cup r_{2}'\right)$.

טענה: $L\left(r'\right)=L^{R}$

הוכחה: כדי להראות $L\left(r'\right)=L^{R}$, נראה כי $w^{R}\in L\left(r'\right)\Longleftrightarrow w\in L$.

$$w\in L\Longleftrightarrow w\in L\left(r\right)=L\left(r_{1}\cup r_{2}\right)\underbrace{=}_{\text{semantics}}L\left(r_{1}\right)\cup L\left(r_{2}\right)\Longleftrightarrow w\in L\left(r_{1}\right)\vee w\in L\left(r_{2}\right)\underbrace{\Longleftrightarrow}_{\text{assumption}}$$

$$w^{R}\in L\left(r_{1}'\right)\vee w^{R}\in L\left(r_{2}'\right)\Longleftrightarrow w^{R}\in L\left(r_{1}'\right)\cup L\left(r_{2}'\right)=L\left(r_{1}'\cup r_{2}'\right)=L\left(r'\right)$$

מקרה 2: כמו במקרה הקודם, לפי הנחה קיימים r_{1}',r_{2}' עבורם מתקיים $L\left(r_{1}'\right)=L\left(r_{1}\right)^{R},L\left(r_{2}'\right)=L\left(r_{2}\right)^{R}$.
נסמן $r'=\left(r_{2}'\cdot r_{1}'\right)$. (נשים לב כי הפכנו את הסדר!)

טענה: $L(r') = L^R$

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

$$w \in L \Longleftrightarrow w \in L(r) = L(r_1 \cdot r_2) \underbrace{=}_{\text{semantics}} L(r_1) \cdot L(r_2) \Longleftrightarrow$$

$$w = w_1 \cdot w_2 : w_1 \in L(r_1) \wedge w_2 \in L(r_2) \quad \underbrace{\Longleftrightarrow}_{\text{assumption+reverse}}$$

$$w^R = w_2^R \cdot w_1^R : w_1 \in L(r_1') \wedge w_2^R \in L(r_2') \Longleftrightarrow w^R \in L(r_2') \cdot L(r_1') = L(r_2' r_1') = L(r')$$

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

טענה: $L(r') = L^R$

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

$$w \in L \Longleftrightarrow w \in L(r) = L(r_1 \cdot r_2) \underbrace{=}_{\text{semantics}} (L(r_1))^* \Longleftrightarrow$$

$$w = w_1 \cdot w_2 \cdot \ldots \cdot w_k : \forall i\,(1 \le i \le k)\; w_i \in L(r_1) \quad \underbrace{\Longleftrightarrow}_{\text{assumption+reverse}}$$

$$w^R = w_k^R \ldots w_1^R, \forall i\,(1 \le i \le k)\; w_i^R \in L(r_1') \Longleftrightarrow w^R \in (L(r_1'))^* = L(r_1'^*) \underbrace{=}_{\text{semantics}} L(r')$$

תרגיל

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

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

תרגילים

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

א) שפת המילים המכילות 10 כתת־מילה.

פתרון: $(1 \cup 0)^* \, 10 \, (1 \cup 0)^*$

ב) שפת המילים המסתיימות ב־ 11.

פתרון: $(1 \cup 0)^* \, 11$

ג) שפת המילים שאורכן זוגי.

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

ד) שפת המילים בהן יש בדיוק 1 פעם אחת.

פתרון: 0^*10^*

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

פתרון: $(ab \cup ba)^*$

חלק V

שיעור 5

7 למת הניפוח לשפות רגולריות

אם L רגולרית, אזי <u>קיים</u> קבוע N כך <u>שלכל</u> $w \in L$ שמקיימת $|w| \geq N$ <u>קיימת</u> חלוקה למילים $x, y, z \in \sum^*$ כך ש: $w = xyz$ ומתקיימים:

1. $y \neq \varepsilon$ ($|y| \geq 1$)
2. $|xy| \leq N$
3. $xy^iz \in L$ לכל $i \geq 0$

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

נשים לב כי הכיוון השני לאו דווקא נכון.

הערה: לימת הניפוח תופיע במבחן.

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

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

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

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

1. $y \neq \varepsilon$
2. $|xy| \leq n$
3. $xy^iz \in L$ לכל $i \geq 0$

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

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

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

כיצד לפתור? נניח בשלילה ש־L רגולרית. יהי n הקבוע שקיומו מובטח מלמת הניפוח. נבחר $w = \underbrace{\qquad}$, $w \in L$, $|w| \geq n$ ויהי פירוק המילה $w = xyz$ כך ש: $|xy| \leq n$, $1 \leq |y|$ (ז"א: $x = $____,$y = $____,$z = $____). נבחר $i = ?$ וע"פ למת הניפוח $xy^iz \in L$. נרצה להראות כי $xy^iz \notin L$ ובכך להגיע לסתירה.

ראשית נדגים בחירה לא טובה של w:

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

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

כעת נדגים בחירה טובה של w:

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

תרגיל:

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

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

עבור $i = 2$ **נקבל:** $|xy^2z| = 2^n + |y|$ (כי $|w| = |xyz| = 2^n$ והוספנו עוד $|y|$) נשים לב כי: $1 \le |y| \le n$ (כי $n \le |xy|$ וגם $|y| \ge 1$) מכאן מקבלים:

$2^n < 2^n + 1 \le |xy^2z| \le 2^n + n < 2^n + 2^n = 2^{n+1}$ כלומר $2^n < |xy^2z| < 2^{n+1}$,
לכן xy^2z איננו מהצורה a^{2^j} עבור $j \in \mathbb{N}$ כלומר $xy^2z \notin L$ והגענו לסתירה.

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

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

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

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

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

נראה דוגמה לשפה לא רגולרית המקיימת את למת הניפוח

$$\sum = \{a, b\},\quad L = \left\{b^ja^p \mid j \ge 0, \text{p is prime}\right\} \cup \{a\}^*$$

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

1. $w = b^ja^p$ עבור $j \ge 1$, p ראשוני , $j + p \ge n$
2. $w = a^k$, עבור $n \le k$ (גם מילים מהצורה b^0a^p כלולות כאן)

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

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

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

חלק VI

שיעור 6

8 מחלקות שקילות ב־$\sum^*$

תהיינה שתי מחרוזות $x, y \in \sum^*$

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

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

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

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

תכונות היחס $\equiv_L$:

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

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

8.1 משפט ($Myhill - Nerod$):

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

תכונות היחס $\equiv_L$ (המשך):

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

דוגמאות

נניח $\sum = \{a,b\}$

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

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

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

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

1. המחלקות זרות בזוגות (אין מילה ב־$\sum^*$ ששייכת ליותר ממחלקה אחת)
2. איחוד המחלקות הוא $\sum^*$
3. לכל שתי מילים מאותה מחלקה <u>אין</u> סיפא מפרידה ביחס ל־L.
4. לכל שתי מילים ממחלקות שונות <u>יש</u> סיפה מפרידה ביחס ל־L.

2)

1. מהן מחלקות השקילות של $\equiv_L$ עבור $L=\{w\mid |w| \bmod 3=0\}$
2. בנה אס"ד מינימלי ל־L.

תשובה: $S_1=\{w\mid |w| \bmod 3=0\}$, $S_2=\{w\mid |w| \bmod 3=1\}$, $S_3=\{w\mid |w| \bmod 3=2\}$ קל לראות (יש להסביר) כי לכל שתי מילים מאותה מחלקה אין סיפא מפרידה ולכל שתי מילים ממחלקות שונות יש סיפא כזו, כמו כן ברור כי $S_1\cup S_2\cup S_3=\sum^*$ (שהן זרות בזוגות).

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

שרטוט במחברת

אצלנו $\delta\left([a],b\right)=\underbrace{[aa]=[ab]}_{\text{same department}}$. שמות המצבים מתייחסים למחלקה כולה, למשל אצלנו $[aa]$ו־$[ab]$ מתייחס לאותו מצב.

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

תשובה: $\begin{array}{ll} S_k = \{w \mid \#a(w) - \#b(w) = k\} & k \geq 0 \\ T_k = \{w \mid \#b(w) - \#a(w) = k\} & k > 0 \end{array}$ נשים לב שכל מילה ב־$\sum^*$ נמצאת בדיוק באחת משתי המחלקות. כלומר, הן זרות בזוגות ואיחודן הוא$\sum^*$.

נראה שלכל 2 מילים מאותה מחלקה אין סיפא מפרידה: לכל $0 < k$ אם $x, y \in S_k$ אזי עבור $z \in T_k$ נקבל: $yz \in L \wedge xz \in L$ לכל z . אחרת, נקבל $xz \notin L \wedge yz \notin L$. לכן, כל המילום ב־$S_k$ עומדות ביחס $\equiv_L$ זו לזו.

הטענה נכונה גם עבור S_0 וגם על T_K, $k > 0$.

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

1. $x \in S_i$, $y \in S_j$ $\qquad i \neq j$
2. $x \in T_i$, $y \in T_j$ $\qquad i \neq j$
3. $x \in S_i$, $y \in T_j$ עבור $j > 0, i \geq 0$

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

מקרה 2: סימטרי למקרה 1.

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

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

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

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

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

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

חלק VII

שיעור 7

9 דקדוקים

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

הגדרה:

1. עבור $\varphi_1, \varphi_2 \in (V \cup \sum)^*$ נאמר כי φ_2 נגזר ישירות מ־φ_1 ונסמן $\varphi_1 \underset{\prod}{\Longrightarrow} \varphi_2$ אם קיימת הפקה $g_i \longrightarrow h_i$ כך ש: $\begin{matrix} \varphi_1 = r \cdot g_i \cdot s \\ r_2 = r \cdot h_i \cdot s \end{matrix}$ כאשר $r, s \in (V \cup \sum)^*$, כלומר ניתן לעבור מ־φ_1 ל־φ_2 ע"י הפעלה אחת של כלל מ־$\prod$
2. נסמן $\varphi_1 \overset{*}{\underset{\prod}{\Longrightarrow}} \varphi_2$ אם ניתן להגיע מ־φ_1 ל־φ_2 ע"י הפעלת מס' סופי של כללים מ־$\prod$ בזה אחר זה.
$$\varphi_1 \underset{\prod}{\Longrightarrow} U_1 \underset{\prod}{\Longrightarrow} \ldots \underset{\prod}{\Longrightarrow} \varphi_2 \in (\sum \cup V)^*$$

שפת הדקדוק: יהי הדקדוק $G = (\sum, V, S, \prod)$, השפה המתקבלת מ־G היא $L(G) = \left\{ w \in \sum^* \mid S \overset{*}{\underset{\prod}{\Longrightarrow}} w \right\}$

9.1 דקדוק רגולרי:

הוא דקדוק שכללי הגזירה ב־$\prod$ הם מהצורה הבאה בלבד: $\begin{matrix} U, W \in V, a \in \sum \\ U \in V, a \in \sum \end{matrix}$ כלומר: $\begin{cases} U \longrightarrow aW \\ U \longrightarrow a \\ U \longrightarrow \varepsilon \end{cases}$

הצד הימני מוגבל לאופציות:

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

דוגמה לדקדוק רגולרי

1. $G_1 : \begin{matrix} S \longrightarrow aB \vert a \\ A \longrightarrow aB \vert a \\ B \longrightarrow bA \end{matrix}$ כאן: $L(G_1) = \{ab\}^* a$
2. דקדוק שקול ל־G_1בעל שני משתנים בלבד: $G_2 : \begin{matrix} S \longrightarrow aB \vert a \\ B \longrightarrow bS \end{matrix}$, מתקיים $L(G_1) = L(G_2)$

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

א) $S \Longrightarrow aB \Longrightarrow abA \Longrightarrow abaB \Longrightarrow ababA \Longrightarrow ababa$

ב) $S \Longrightarrow aB \Longrightarrow abS \Longrightarrow abaB \Longrightarrow ababS \Longrightarrow ababa$

דוגמה לדקדוק לא רגולרי

$S \longrightarrow aS\underline{a}|bS\underline{b}|\underline{SS}|a|b|\varepsilon|\underline{ab}|aS$ (המסומן בקו תחתון הוא מה שלא עונה להגדרה של דקדוק רגולרי)

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

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

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

10 שפות חסרות־הקשר

10.1 דקדוק חסר הקשר:

זהו דקדוק בו כל כללי הגזירה של $\prod$ הם מהצורה $X \longrightarrow \beta$ כאשר $X \in V$ ו־$\beta \in (\sum \cup V)^*$ (הערה: הכלל $X \in V, X \longrightarrow \varepsilon$ מותר)

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

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

דוגמאות לדקדוק חסר הקשר

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

10.2 עץ גזירה עבור דקדוק ח"ה

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

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

תרגיל

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

תשובה

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

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

1. שפת המילים המתחילות ברצף ab
2. שפת המילים באורך זוגי
3. שפת המילים באורך אי־זוגי
4. $L = \{w \mid \#b(w)\, mod2 = 0\}$

תשובות:

1. $$\begin{array}{c} S \longrightarrow abX \\ X \longrightarrow aX|bX|\varepsilon \end{array}$$
2. $S \longrightarrow abS|baS|aaS|bbS|\varepsilon$ או $\begin{array}{c} S \longrightarrow AAS|\varepsilon \\ A \longrightarrow a|b \end{array}$
3. עבור אי זוגי נחליף את הכלל $S \longrightarrow \varepsilon$ בתשובות ל־2 ב־$S \longrightarrow a|b$
4. $$\begin{array}{c} S \longrightarrow AbAbS|A \\ A \longrightarrow aA|\varepsilon \end{array}$$

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

$$\begin{array}{c} S \longrightarrow aB|cB|cA|a|b|\varepsilon \\ A \longrightarrow aA|\varepsilon|b \\ B \longrightarrow bB|aB|cB|\varepsilon \end{array}$$

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

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

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

תשובות:

1. $\begin{array}{l} S \longrightarrow aSb|aA|bB \\ A \longrightarrow aA|\varepsilon \\ B \longrightarrow bB|\varepsilon \end{array}$ נראה לדוגמה גזירה עבור $w = aaabb$: $S \longrightarrow aSb \longrightarrow aaSbb \longrightarrow aaaAbb \longrightarrow aaabb$

2. $S \longrightarrow aaaSb|aaSb|aSb|Sb|\varepsilon$

חלק VIII

שיעור 8

11 הצורה הנורמלית של חומסקי

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

$\begin{array}{lll} V_0, V_1, V_2 \in V & , & V_0 \longrightarrow V_1V_2 \\ a \in \sum & , & V_0 \longrightarrow a \end{array}$. דקדוק כזה נקרא דקדוק בצורה הנורמלית של חומסקי.

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

כלומר אין ב־G'' כללים מהצורה: $\begin{array}{ll} X \longrightarrow \varepsilon, X \in V & \text{(חיובי)} \\ X \longrightarrow Y, X, Y \in V & \text{(מסתעף)} \end{array}$ כך ש: $L(G) = L(G') \cup \{\varepsilon\}$

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

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

אם $a \in \sum$, $A \longrightarrow a$ כלל ב־G' אז הוא גם כלל ב־G''. (כלל = הפקה). נותרו כללים מהצורה $(*) := A \longrightarrow$ $X_1X_2 \ldots X_t$ כאשר $2 \leq t$ ו־$X_i \in \sum \cup V$ לכל $1 \leq i \leq t$.

עבור כל כלל מהצורה הנ"ל: $(*)$וטרמינל $a \in \sum$ (למשל $X_i = a$) נוסיף משתנה S_a ל־V ואת הכלל $S_a \longrightarrow a$ ל־G'. נחליף את כל המופעים של a ב־$(*)$ ב־S_a. הכללים הנותרים שאינם מצורה זו הם: $A \longrightarrow X_1X_2 \ldots X_t$ כאשר $3 \leq t$ ו־ $\left(\bigcup_{a \in \sum} \{S_a\}\right)$ $X_i \in V \cup$ לכל $1 \leq i \leq t$. עבור כל כלל כזה נוסיף משתנים חדשים (שיהיו שונים מכלל לכלל) $Y_1, Y_2, \ldots, Y_{t-2}$, ונוסיף במקומו את הכללים הבאים ל־G'':

$$\begin{array}{c} A \longrightarrow X_1Y_1 \\ Y_1 \longrightarrow X_2Y_2 \\ \vdots \\ Y_{t-3} \longrightarrow X_{t-2}Y_{t-2} \\ Y_{t-2} \longrightarrow X_{t-1}X_t \end{array}$$

דוגמה:

יהי G' דקדוק ח"ה בעל הכללים הבאים: $\begin{array}{l} S \longrightarrow Aa|a \\ A \longrightarrow Ab|b|BBAa|BAa \\ B \longrightarrow AS|b \end{array}$ נעביר אותו לדקדוק ח"ה בצורה הנורמלית של חומסקי ע"י האלגוריתם: (נשים לב שהדקדוק חיובי ומסתעף!)

נשאיר את הכללים: $\begin{matrix} S\longrightarrow a \\ A\longrightarrow b \\ B\longrightarrow AS|b \end{matrix}$ ועבור הכללים שנותרו נחליף את a ב־S_a ואת b ב־S_b כאשר S_b, S_a הם משתנים חדשים שעבורם הוספנו את הכללים: $\begin{matrix} S_a\longrightarrow a \\ S_b\longrightarrow b \end{matrix}$ ונקבל סה"כ : $\begin{matrix} S\longrightarrow AS_a|a \\ A\longrightarrow AS_b|b|BBAS_a|BAS_a \\ B\longrightarrow AS|b \\ S_a\longrightarrow a \\ S_b\longrightarrow b \end{matrix}$

כעת במקום הכלל $A\longrightarrow BBAS_a$ (הוא בעייתי כי אינו עונה על ההגדרה) נוסיף את הכללים: $\begin{matrix} A\longrightarrow BY_1 \\ Y_1\longrightarrow BY_2 \\ Y_2\longrightarrow AS_a \end{matrix}$

(הוספנו משתנים חדשים Y_1, Y_2) ובמקום הכלל $A\longrightarrow BAS_a$, נוסיף

את הכללים: $\begin{matrix} A\longrightarrow BY_3 \\ Y_3\longrightarrow AS_a \end{matrix}$ נשים לב שהוספנו משתנה חדש Y_3 עם האינדקס 3 כי השמות Y_1, Y_2 תפוסים.

כלומר, לכל כלל נגדיר משתנים חדשים שלא תפוסים.

12 למת הניפוח לשפות חסרות־הקשר (משפט בר־הלל)

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

1. $vy\neq\varepsilon$
2. $|vxy|\leq k$
3. $uv^nxy^nz\in L$ לכל $0\leq n$

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

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

תרגיל

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

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

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

נבחן את כל האפשרויות:

1. vxy לא מכילה את האות a.

2. vxy לא מכילה את האות c.

3. vxy לא מכילה את האות b.

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

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

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

תרגיל

נתונה השפה הבאה מעל $\sum = \{a, b, c\}$: $L = \left\{X_1 C X_2 C \ldots C X_k \,|\, k \geq 2 \text{ each } X_i \in \{a, b\}^+ \text{ and for some } i \neq j, \quad X_i = X_j\right\}$
הוכיחו כי L אינה ח"ה.

פתרון: נניח בשלילה כי L ח"ה. נגדיר את השפה הבאה: $\underbrace{L' = L \cap L\left(\underbrace{a^* b^* c a^* b^*}_{\text{Regular Expression}}\right)}_{\text{Regular Language}}$ אזי, מסגירות תחת חיתוך עם שפה רגולרית נובע כי גם L' חסרת הקשר. נראה כי L' אינה ח"ה , נגיע לסתירה ונסיק ש־L אינה ח"ה.

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

1. x מכילה את התו c

2. x לא מכילה את התו c

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

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

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

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

חלק IX

שיעור 9

13 מעבר מדקדוק חסר־הקשר חיובי לא מסתעף לדקדוק חסר־הקשר חיובי ומסתעף

נתון דקדוק חסר הקשר חיובי לא מסתעף G', כלומר בדקדוק יש כללים מהצורה $\begin{matrix}A\longrightarrow B\\ A,B\in V\end{matrix}$ (נקרא גם כלל יחידה) ואין לנו כללי ε.

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

יהי G' דקדוק חסר הקשר שכללי הגזירה שלו הם: $\begin{aligned}S&\longrightarrow Q|b\\ Q&\longrightarrow R|a\\ R&\longrightarrow ba|bRa\end{aligned}$

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

בשלב הראשון: ניקח את הכללים שאינם כללי יחידה ל־G'': $\begin{aligned}S&\longrightarrow b\\ Q&\longrightarrow a\\ R&\longrightarrow ba|bRa\end{aligned}$

בשלב השני: נמצא את כל זוגות המשתנים A,B, $A\neq B$ המקיימים: $A\overset{*}{\underset{G'}{\Longrightarrow}}B$ (במקרה שלנו, $\begin{aligned}S&\longrightarrow Q\\ Q&\longrightarrow R\end{aligned}$

לכן: $\begin{aligned}S&\overset{*}{\underset{G'}{\Longrightarrow}}Q\\ Q&\overset{*}{\underset{G'}{\Longrightarrow}}R\end{aligned}$ וגם $S\overset{*}{\underset{G'}{\Longrightarrow}}R$, מטרנזיטיביות)

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

1) מכיוון ש: $S\overset{*}{\underset{G'}{\Longrightarrow}}R$ ו־ $S\overset{*}{\underset{G'}{\Longrightarrow}}Q$, נצרף לדקדוק השקול כל כלל מהצורה: $S\longrightarrow\alpha$, כאשר $Q\longrightarrow\alpha$ הוא כלל ב־G' שאינו כלל יחידה או $R\longrightarrow\alpha$ הוא כלל ב־G' שאינו כלל יחידה.

2) מכיוון ש: $R \overset{*}{\underset{G'}{\Longrightarrow}} Q$ נצרף לדקדוק השקול כל כלל מהצורה $Q \longrightarrow \alpha$ כאשר $R \longrightarrow \alpha$ הוא כלל ב־G' שאינו כלל יחידה. (מ־1 ו־2 נובע כי במקרה שלנו, יכנסו הכללים: **מ־1:** $S \longrightarrow a|ba|bRa$ **מ־2:** $Q \longrightarrow ba|bRa$ במקום הכללים: $\begin{matrix} S \longrightarrow Q \\ Q \longrightarrow R \end{matrix}$)

סה"כ קיבלנו דקדוק חסר הקשר חיובי ומסתעף שקול, G'', שכללי הגזירה שלו הם: $\begin{matrix} S \longrightarrow b|a|ba|bRa \\ Q \longrightarrow a|ba|bRa \\ R \longrightarrow ba|bRa \end{matrix}$ (ומכיוון שאי אפשר לעבור ל־Q, אפשר להשמיטו: $\begin{matrix} S \longrightarrow b|a|ba|bRa \\ R \longrightarrow ba|bRa \end{matrix}$)

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

13.1 הקשר בין דקדוק בצורה הנורמלית של חומסקי ללמת הניפוח לשפות חסרות־הקשר

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

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

14 אוטומט מחסנית:

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

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

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

[2] יכול להופיע במבחן שאלות דומות

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

השפה המתקבלת ע"י אוטומט מחסנית:

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

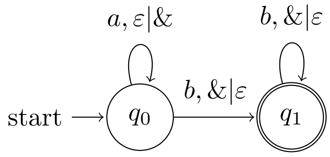
שימו לב: המודל הנ"ל הוא א־דטרמיניסטי. המודל הדטרמיניסטי לא שקול. כאשר מוזכר בקורס אוטומט מחסנית, נתכוון למודל הא־דטרמיניסטי (אלא אם צויין במפורש אחרת).

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

דוגמה

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

$b,\&|\varepsilon$



באופן כללי־ סימונים במעברים בין מצב למצב:

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

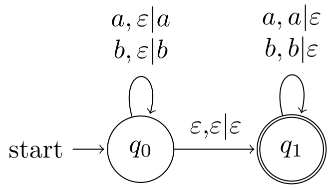
אם $\alpha=\varepsilon$ ־ מעבר ללא קריאת תו (תיתכן תלות במחסנית).

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

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

דוגמה

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



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

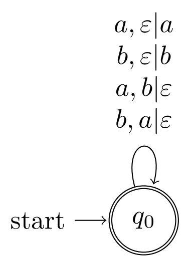
נשים לב שהמחסנית נותנת לנו להשוות מילה והיפוכה. לעומת זאת, השפה $L=\left\{ww\left|w\in\sum^{*}\right.\right\}$ לא ח"ה (חשבו מדוע אי אפשר לבנות לה אוטומט מחסנית).

הראינו ע"י לימת הניפוח לשפות ח"ה כי השפה $L=\left\{a^{i}b^{i}c^{I}|i\geq0\right\}$ לא ח"ה. כעת יש לנו אינטואיציה לכך, משום שאנו רואים שלא מספיקה כאן מחסנית אחת, מכיוון שלאחר ההשוואה בין ה־a־ים ל־b־ים, המחסנית תהיה ריקה, ולא נדע להשוות בין ה־c־ים ל־b־ים.

הערה: זו איננה נחשבת להוכחה!

דוגמה נוספת:

$$L=\left\{w\mid\#a\left(w\right)=\#b\left(w\right)\right\} \quad \sum=\left\{a,b\right\}$$



אינטואיציה:

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

חלק X

שיעור 10

15 הוכחות נכונות

ראינו כיצד להוכיח נכונות של ביטויים רגולריים[3].

15.1 כיצד נוכיח נכונות של אס"ד או דקדוק?

נראה דוגמאות:

15.1.1 דוגמה לתרגיל עם הוכחת נכונות של אס"ד:

תהי $L=\left\{w\mid\left|w\right|\mod7=4\right\}$ מעל הא"ב $\sum=\left\{a,b\right\}$

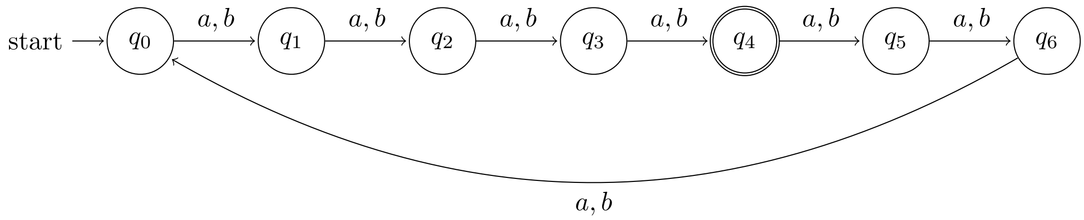
[3] בתרגיל 4, כאשר נתבקשנו להוכיח שאם L רגולרית אזי L^R רגולרית.

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

כלומר: צריך להוכיח כי $L(M)=L$

פתרון: נבנה אוטומט M:

$$Q=\{q_i \mid i=0,\ldots,6\}$$
$$F=\{q_4\}$$
$$\forall\sigma\in\sum,\forall q_i\in Q:\ \delta(q_i,\sigma)=q_{(i+1)\mod 7}$$



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

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

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

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

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

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

נתבונן באות $\sigma\in\sum$. עבורה מתקיים:

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

ולכן ע"פ הנחת האינדוקציה על w_1, מתקיים: $\delta^*(q_0,w_1)=q_{(i-1)\mod 7}$. בנוסף, ע"פ בניית M מתקיים $\delta\left(q_{(i-1)},\sigma\right)=q_i$, וסה"כ קיבלנו: $\delta^*\left(q_0,\underbrace{w_1\sigma}_{w}\right)=\delta\left(q_{(i-1)},\sigma\right)=q_i$ כנדרש.

הוכחת $L(M)=L$:

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

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

15.1.2 דוגמה להוכחת נכונות דקדוק:

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

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

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

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

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

צעד האינדוקציה: נחלק שלב זה לשלושה מקרים: w התקבלה כך שכלל הגזירה הראשון שהופעל היה:
1) $S \longrightarrow aSb$
2) $S \longrightarrow bSa$
3) $S \longrightarrow SS$

נבחן את המקרים השונים:

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

2) ההוכחה באופן דומה למקרה א'.

3) w היא מהצורה $w = y_1y_2$, כך ש: $y_1, y_2 \in L(G)$. שתי מילים שהתקבלו ע"י סדרת גזירות באורך קטן מ־n. לפי הנחת האינדוקציה על y_1, y_2 מתקיים $y_1 \in L \Longrightarrow \#a(y_1) = \#b(y_1)$ וגם $y_2 \in L \Longrightarrow \#a(y_2) = \#b(y_2)$ עבור המילה w מתקיים $\#a(w) = \#a(y_1) + \#a(y_2)$ וגם $\#b(w) = \#b(y_1) + \#b(y_2)$, מכאן מתקיים: $\#a(w) = \#b(w) \Longrightarrow w \in L$.

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

16 חזרה ־ אוטומט מחסנית:

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

פתרון: נשים לב שאם $w \in L$ אז $w = x\#y$ ומתקיים אחד מהתנאים הבאים:

1) $|x| < |y|$
2) $|x| > |y|$
3) $|x| = |y|$

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

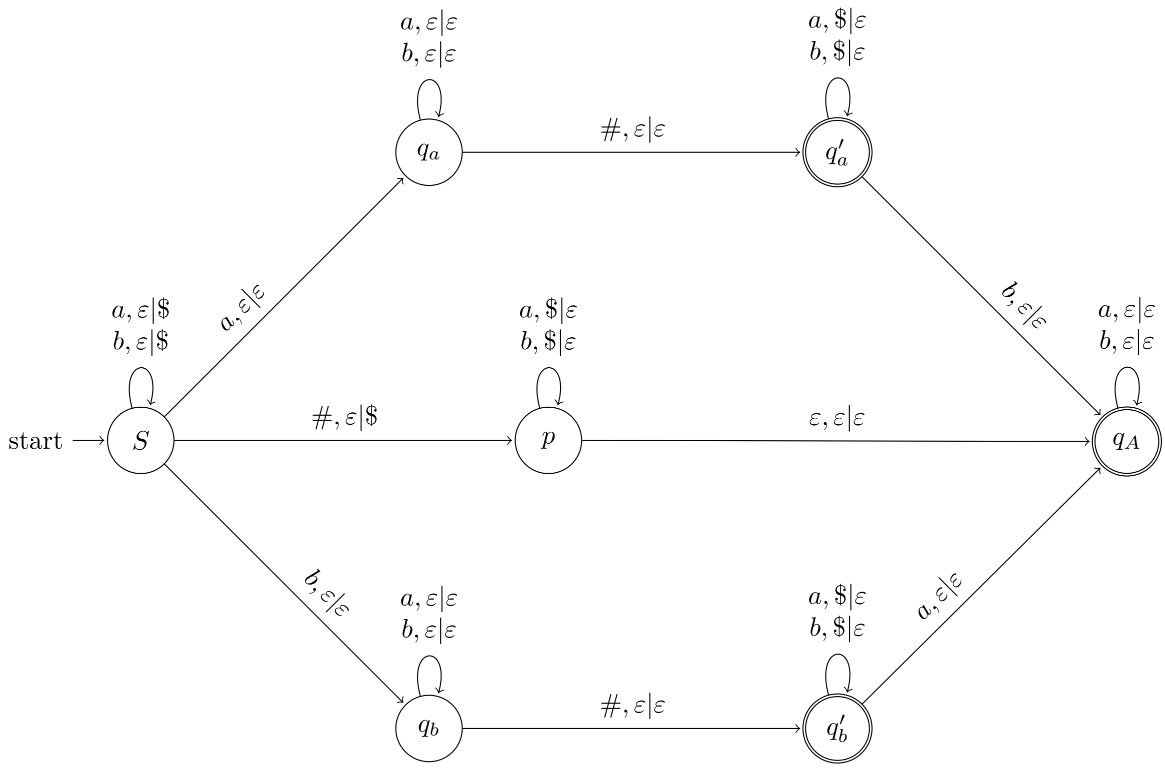
היינו רוצים לבנות אמל"ד לכל אחד מהמקרים ולאחד את שלושתם לאמל"ד יחיד (אפשר, כי שפות ח"ה סגורות לאיחוד) .

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

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

מקרה 1 - $|x| < |y|$

מקרה 2 - $|x| > |y|$ איחוד כל המקרים (וחיסכון ע"י ויתור על q ו־q') נותן:



חלק XI

שיעור 11

17 שפות תלויות הקשר Context-Sensitive languages

17.1 דקדוק תלוי הקשר

הגדרה: דקדוק תלוי הקשר הינו דקדוק שבו כל ההפקות הן מהצורה $U \longrightarrow V \,:\, u,v \in \left(\sum \cup V\right)^{+}$ כאשר $|u| \le |v|$.

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

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

משפט: שפה L תלוית הקשר $\Longleftrightarrow$ קיים דקדוק תלוי הקשר G כך ש-$L - \{\varepsilon\} = L(G)$ [4]

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

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

דוגמה לשפה תלוית הקשר שאינה חסרת הקשר:

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

$G:$

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

נראה דוגמה לגזירה של $w = aabbcc$: נראה $S \overset{*}{\underset{G}{\Longrightarrow}} w$:

$$S \longrightarrow a\underline{S}BC \longrightarrow a\underline{aB}CBC \longrightarrow aab\underline{CB}C \longrightarrow aa\underline{bB}CC \longrightarrow aabb\underline{CC} \longrightarrow aabbc\underline{C} \longrightarrow aabbcc$$

[4] זאת, כדי להרשות ε בשפה L

17.2 סגירויות של שפות תלויות הקשר

משפחת השפות תלויות ההקשר סגורה תחת הפעולות: איחוד, חיתוך, משלים, שרשור, $*$־קליני.[5] סגירות תחת השלמה הוכח רק בסוף שנות ה־80 .

18 מכונות טיורינג Turing Machine

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

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

מכונת טיורינג מורכבת מהרכיבים הבאים:

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

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

Q־ קבוצת מצבים

$\sum$ ־ א"ב מעליו אנו עובדים

q_0 ־ מצב התחלתי

$\sum \cup \{\hbar\} \leq \Gamma$

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

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

עבור מכונת טיורינג דטרמיניסטית:

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

עבור מכונת טיורינג א־דטרמיניסטית:

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

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

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

[5] חשוב לדעת למבחן!

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

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

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

19 מכונת טיורינג א־דטרמיניסטית חסומה לינארית[6]

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

+	TURING	+

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

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

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

19.1 שאלה פתוחה:

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

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

תרגיל

תנו תיאור מילולי וציירו דיאגרמה של LBA המקבל את $L = \{a^n b^n c^n \mid n \geq 0\}$

אופן פעולת המכונה על קלט w:

1. כל עוד התו הוא x[7] המכונה מזיזה את הראש ימינה.
2. אם התו הנקרא הוא $\hbar \Longleftarrow$ מקבלת.
3. אם התו הנקרא הוא b או $c \Longleftarrow$ דחה.
4. המכונה מוחקת תו a ראשון ע"י $\hbar$ ומזיזה את הראש ימינה.

[6] נקרא גם אוטומט חסום לינארית ־ LBA

[7] x הינו תו שלא נמצא בקלט, נמצא בא"ב הסרט ויסמן לנו תוים שנמחקו מהקלט במהלך הריצה.

5. כל עוד התו הוא a או x המכונה מזיזה את הראש ימינה.
   (א) אם התו הנקרא הוא c או $\hbar$ $\Longleftarrow$דחה.
   (ב) אם התו הנקרא הוא b, המכונה משכתבת אותו להיות x.
6. כל עוד התו הנקרא הוא b או x המכונה מזיזה את הראש ימינה.
   (א) אם התו הנקרא הוא a או $\hbar$ $\Longleftarrow$דחה.
   (ב) אם התו הנקרא הוא c המכונה משכתבת אותו להיות x.
7. כל עוד התו הנקרא הוא c המכונה מזיזה את הראש ימינה.
   (א) אם התו הנקרא הוא a או b $\Longleftarrow$דחה.
   (ב) אם התו הנקרא הוא $\hbar$, המכונה מזיזה את הראש שמאלה עד לתו הראשון שמימין לרצף הרווחים שבתחילת הסרט.
8. המכונה חוזרת לשלב 1.

משמעות המצבים: אופן פעולת המכונה הוא איטרטיבי.

קלט הלולאה: הינו המילה הכתובה מייד לאחר רצף ה־$\hbar$ השמאלי (אם יש) ועד תחילת תו ה־$\hbar$ מימין לקלט (סוף הסרט)

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

האינווריאנטה (תכונה נשמרת)

התכונות הנשמרות מאיטרציה לאיטרציה:

1. הקלט המקורי הוא מהצורה $a^*b^*c^*$ וקלט הלולאה הוא מהצורה: $a^*x^ib^*x^ic^*$
2. מספר התווים המחוקים (בין אם ע"י $\hbar$ או x) מרצף ה־a־ים שווה למס' התווים המחוקים מרצף הb־ים וה־c־ים.

מגיעים למצב מקבל אם קיימת ריצה של לולאה שבסיומה מחוקים כל תווי ה־ a , b ו־c והקלט הוא מהצורה $a^nb^nc^n$ עבור $0 \leq n$.

שאלה: נתונה השפה הבאה: $L = \left\{w\#w \,|\, w \in \{0,1\}^*\right\}$

הסבירו במילים את אלגוריתם הפעולה ש־LBA מקבל אותה.

פתרון:

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

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

פתרון: אופן פעולת המכונה על קלט w:

1. המכונה בודקת אם הקלט הוא מהצורה $\{0,1\}^* \# \{0,1\}^*$ (המכונה תזיז את הראש ימינה כל עוד היא קוראת 0 או 1. אם היא הגיעה לסוף הסרט מבלי לקרוא $\#$ או לקרוא יותר מ־# אחת, דחה)
2. אם התו הנקרא הוא $\#$ המכונה בודקת כי כל הסימנים מימין לו נמחקו, אם כן נמחקו $\Longleftarrow$ המכונה מקבלת.
   (א) אחרת דחה.
3. יהי σ התו הנקרא במילה.
   (א) המכונה זוכרת את σ וכותבת במקומו x.
   (ב) המכונה מזיזה את הראש לתו הראשון ששונה מ־x מימין ל־$\#$ ובודקת אם שווה ל־σ
      i. אם לא $\Longleftarrow$דחה.
      ii. אחרת, המכונה כותבת במקומו x.
   (ג) המכונה מזיזה את הראש שמאלה עד ה־x הראשון משמאל ל־$\#$
4. המכונה מזיזה את הראש צעד אחד ימינה
5. חוזרת למצב 2.

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

חלק XII

הוכחות של משפטים מרכזיים

ההוכחות בחלק זה מתבססות על הרצאות של פרופ' משה קופל.

20 לימת הניפוח לשפות רגולריות - Pumping Lemma

יהי M אס"ד בעל m מצבים ותהי x מילה השייכת ל-$x \in L(M)$ כך ש-$m \leq |x|$. אזי ניתן לפרק את x ל-uvw כך ש:

1) $|uv| \leq m$ (האורך של uv קטן שווה ל-m)

2) $|v| > 0$

3) לכל $i = 0, 1, \ldots$ $uv^iw \in L(M)$

הוכחה: תוך כדי אכילת m האותיות הראשונות של x, אנו נתקלים ב-$m+1$ מצבים (לא בהכרח ייחודים). מכיוון שיש ב-M רק m מצבים, אנו בהכרח נתקלים בלפחות מצב אחד, שייקרא q, לפחות פעמיים. נקרא לרישא של x, עד למופע הראשון של q, באות u, ונקרא לקטע של x בין ההופעה הראשונה של q להופעה השנייה שלו, באות v. נשים לב ש-$|uv| \leq m$ וכמו כן, $|v| > 0$. אנו יודעים כי $\delta^*(q_0, u) = q$, $\delta^*(q, v) = q$, $\delta^*(q, w) \in F$ ולכן: $\delta^*\left(q_0, uv^iw\right) =$

$\delta^*\left(q, uv^iw\right) = \delta^*(q, w) \in F$ (ומכאן ברור כי $uv^iw \in L(M)$)

כל שפה רגולרית מקיימת את לימת הניפוח.

21 תכנות הסגור של שפות רגולריות

21.1 סגירות לאיחוד

משפט: תהיינה L_1, L_2 שפות רגולריות. אזי, $L_1 \cup L_2$ היא רגולרית.

הוכחה: יהי M_1 אס"ד $\left\{\sum_1, Q_1, q_0^1, F_1, \delta_1\right\}$ כך ש- $L(M_1) = L_1$ ויהי M_2 אס"ד $\left\{\sum_2, Q_2, q_0^2, F_2, \delta_2\right\}$ כך ש- $L(M_2) = L_2$.

צ"ל: קיים אסל"ד M $\left\{\sum, Q, q_0, F, \delta\right\}$כך ש-$L(M) = L_1 \cup L_2$ (נוח יותר לבנות אסל"ד וכבר הראינו כי מאסל"ד ניתן לבנות אס"ד)

נגדיר את M כך:

$$\sum = \sum_1 \cup \sum_2$$
$$Q = Q_1 \cup Q_2 \cup (q_0)$$
$$q_0$$
$$F = F_1 \cup F_2$$

$$\delta(q, a) = \begin{cases} \delta_1(q, a) & q \in Q_1 \\ \delta_2(q, a) & q \in Q_2 \end{cases}$$
$$\delta(q_0, a) = \emptyset$$
$$\delta(q, \varepsilon) = \left\{q_0^1, q_0^2\right\}$$

(כאשר $\delta(q_0, a) = \emptyset$ מציין כי אין צלע המובילה מ-q_0 שמקבלת את a)

21.2 סגירות להשלמה

משפט: תהי L שפה רגולרית, אזי $\overline{L}$ רגולרית.

הוכחה: יהי M אס"ד $\left\{\sum, Q, q_0, F, \delta\right\}$כך ש-$L\left(M\right)=L$. נבנה אס"ד $\overline{M}$ כך ש-$L\left(\overline{M}\right)=\overline{L}$. נגדיר את $\overline{M}$ כך: $\left\{\sum, Q, q_0, \overline{F}=Q\backslash F, \delta\right\}$

21.3 סגירות לחיתוך

משפט: תהיינה L_1, L_2 שפות רגולריות. אזי, $L_1\cap L_2$ היא רגולרית.

הוכחה: $L_1\cap L_2=\overline{\overline{L_1}\cup\overline{L_2}}$ (אגף ימין רגולרי, הוכחנו במשפטים הקודמים, ולכן גם אגף שמאל, ע"פ חוקי דה-מורגן)

21.4 סגירות לשרשור

משפט: תהיינה L_1, L_2 שפות רגולריות. אזי, $L_1\cdot L_2$ היא רגולרית.

הוכחה: יהי M_1 אס"ד $\left\{\sum_1, Q_1, q_0^1, F_1, \delta_1\right\}$ כך ש- $L\left(M_1\right)=L_1$ ויהי M_2 אס"ד $\left\{\sum_2, Q_2, q_0^2, F_2, \delta_2\right\}$ כך ש- $L\left(M_2\right)=L_2$, ונניח גם כי $Q_1\cap Q_2=\emptyset$ (מכיוון שאנחנו בוחרים את שמות המצבים, ניתן לדרוש תנאי זה)

צ"ל: קיים אסל"ד M $\left\{\sum, Q, q_0, F, \delta\right\}$כך ש-$L\left(M\right)=L_1\cdot L_2$. נגדיר את M כך:

$$\begin{gathered}\sum=\sum{}_1\cup\sum{}_2\\ Q=Q_1\cup Q_2\cup\left(q_0\right)\\ q_0=q_0^1\\ F=F_2\end{gathered}$$

$$\delta\left(q,a\right)=\begin{cases}\delta_1\left(q,a\right) & q\in Q_1\\ \delta_2\left(q,a\right) & q\in Q_2\end{cases}$$

$$\delta\left(q,\varepsilon\right)=\left\{q_0^2\right\}\quad\left(q\in F_1\right)$$

21.5 סגירות ל- * קליני

משפט: תהי L שפה רגולרית. אזי, L^* היא רגולרית.

הוכחה: יהי M אס"ד $\left\{\sum, Q, q_0, F, \delta\right\}$ כך ש- $L\left(M\right)=L$

צ"ל: קיים אסל"ד M' $\left\{\sum', Q', q_0', F', \delta'\right\}$כך ש-$L\left(M'\right)=L^*$.

נגדיר את M' כך:

$$\begin{gathered}\sum=\sum{}'\\ Q=Q'\cup\left\{q_0'\right\}\\ F'=\begin{cases}F\cup\left\{\hat{q_0}\right\} & q_0\in F\\ F\cup\left\{q_0\right\} & q_0\notin F\end{cases}\\ q_0'=q_0\end{gathered}$$

$$\delta'\left(q,a\right)=\begin{cases}\delta\left(q,a\right) & q\in Q\\ \delta\left(q_0,a\right) & q=\hat{q_0}\end{cases}$$

$$\delta'\left(q,\varepsilon\right)=\left\{\hat{q_0}\right\}\quad\left(q\in F\right)$$

22 משפט קליני ($Kleene$):

השפה L היא רגולרית אםם קיים ביטוי רגולרי המייצג את L.

הוכחה:

$(\Longrightarrow)$ טריוויאלי. נובע מסגירות הפעולות: שרשור, איחוד, כוכבית...

$(\Longleftarrow)$ תהי L שפה רגולרית ויהי $M=\left\{\sum,\underbrace{Q}_{\{q_1,\ldots,q_m\}},q_1,F,\delta\right\}$ אס"ד כך ש-$L=L\left(M\right)$.

צ"ל שקיים ביטוי α כך ש-α מייצג את L.

נגדיר ביטוי $R_{i,j}^{k}$ כך:

$$R_{i,j}^{k}=\left\{\underbrace{w}_{s_1,\ldots,s_l}\mid\delta\left(q_i,s_1\right)=q_{i_1},\delta\left(q_{i_1},s_2\right)=q_{i_2},\ldots,\delta\left(q_{i_{l-1}},s_l\right)=q_j\ \wedge\ i_1,i_2,\ldots,i_{l-1}\leq k\right\}$$

טענה: לכל i,j,k ניתן לייצג את $R_{i,j}^{k}$ ע"י ביטוי רגולרי.

הוכחת הטענה: באינדוקציה על k: עבור $k=0$ מתקיים: $R_{i,j}^{0}\subseteq\sum$

עבור $k+1$: נניח באינדוקציה שלכל i,j מתקיים ש- $R_{i,j}^{k}$ הוא ביטוי רגולרי. צ"ל ש-$R_{i,j}^{k+1}$ הוא ביטוי רגולרי.

$$R_{i,j}^{k+1}=R_{i,j}^{k}\cup R_{i,k+1}^{k}\cdot\left(R_{k+1,k+1}^{k}\right)^{*}\cdot R_{k+1,j}^{k}$$

כעת נתאר:

$$L\left(M\right)=\bigcup_{q_j\in F}R_{1,j}^{m}$$

■

הרעיון הוא שכל $R_{1,j}^{m}$ מתאר מצב שהתחלנו באחד וסיימנו במצב מקבל j כלשהו.

23 משפט ($Myhill-Nerode$)

הגדרה: תהי $L\subseteq\sum^{*}$ שפה כלשהי ותהינה $x,y\in\sum^{*}$ מילים. נאמר ש: $x\equiv_{L}y$ (x שקול ל-y ב-L) אם לכל $w\in\sum^{*}$ מתקיים $xw\in L\Longleftrightarrow yw\in L$.

הגדרה: הקבוצה $S\subseteq\sum^*$ תקרא קבוצה פורשת ל־L אם לכל $w\in\sum^*$ קיימת מילה $x_i\in S$ כך שמתקיים: $x_i\equiv_L w$

משפט $(Myhill-Nerode)$

השפה L רגולרית אסם קיימת עבורה קבוצה פורשת סופית.

הוכחה: $(\Longleftarrow)$ נניח כי L רגולרית, צ"ל קיימת עבורה קבוצה פורשת סופית. מכיוון ש־L רגולרית, יהי M אס"ד המקבל את L המוגדר כך: $M=\left\{\sum,\underbrace{Q}_{\{q_1,\dots,q_n\}},F,q_1,\delta\right\}$ כך שמתקיים: $L=L(M)$. תהי x_i מילה כך ש: $\delta^*(q_1,x_i)=q_i$. תהי $S=\{x_1,\dots,x_n\}$

טענה: S היא קבוצה פורשת סופית עבור L. צ"ל שעבור כל $y\in\sum^*$ קיימת מילה $x_i\in S$ כך ש: $x_i\equiv y$.
יהי $\delta^*(q_1,y)=q_j$ אזי $x_j\equiv_L y$ ומתקיים:

$$x_jw\in L\Longleftrightarrow\delta^*(q_1,x_jw)\in F\Longleftrightarrow\delta^*(q_j,w)\in F\Longleftrightarrow\delta^*(q_1,yw)\in F\Longleftrightarrow yw\in L$$

$(\Longrightarrow)$ תהי S קבוצה פורשת עבור L, $S=\{\varepsilon=x_1,\dots,x_n\}$. צ"ל קיים אס"ד M כך ש: $L=L(M)$.
יהיו מצבים כמו בהוכחת הכיוון השני: $\sum=\sum$, $Q=\{q_1,\dots,q_n\}$, q_1 , $F=\{q_i\,|\,x_i\in L\}$,
ונגדיר $\delta(q_i,a)=q_j$ כך ש: $x_i\cdot a\equiv_L x_j$. (אז הרחבה של δ לקבל מילה w תהיה: $\delta^*(q_i,w)=q_j$ כך ש: $x_i\cdot w\equiv_L x_j$)

$$w\in L\Longleftrightarrow\varepsilon w\in L\Longleftrightarrow x_1w\in L\Longleftrightarrow\overbrace{x_1w}^{x_1w\equiv_L x_j}\in L\Longleftrightarrow x_j\in L\Longleftrightarrow q_j\in F\Longleftrightarrow$$

$$\delta^*(q_1,w)=q_j\Longleftrightarrow\delta^*(q_1,w)\in F\Longleftrightarrow w\in L(M)$$

24 L היא רגולרית אס"ם קיים דקדוק רגולרי Γ שיוצר אותה

הגדרה:

הדקדוק Γ קרוי דקדוק רגולרי אם כל ההפקות ב־Γ הן מהצורות: $\begin{matrix}X,Y\in V & X\longrightarrow aY\\ a\in\sum & X\longrightarrow a\end{matrix}$. כעת ננסח משפט:

משפט: השפה L היא רגולרית אסם קיים דקדוק רגולרי Γ כך ש: $L=L(\Gamma)$ (ומתקיים: $L=L(\Gamma)\cup\{\varepsilon\}$ או $L=L(\Gamma)$)

משפט: תהי L שפה רגולרית, אזי קיים דקדוק רגולרי Γ כך ש: $L=L(\Gamma)$

הוכחה: יהי $M=\left\{\sum,\underbrace{Q}_{\{q_1,\ldots,q_n\}},F,q_1,\delta\right\}$ אס"ד כך ש: $L=L(M)$ צ"ל קיים דקדוק רגולרי Γ כך ש: $L(\Gamma)=L(M)$

$$\begin{gathered}\sum=\sum\\ V=\{S=X_1,\ldots,X_n\}\\ \begin{cases}\delta(q_i,a)=q_j\Longrightarrow\quad X_i\longrightarrow aX_j\in\prod\\ \delta(q_i,a)\in F\Longrightarrow\quad X_i\longrightarrow a\in\prod\end{cases}\end{gathered}$$

טענה: $L(M)=L(\Gamma)$

טענה 1 (כיוון ראשון): $w\in L(M)\Longrightarrow w\in L(\Gamma)$

תהי $w=a_1\ldots_l\in L(M)$ אזי $\delta^*(q_1,w)\in F$ (w , המילה המתקבלת ע"י האוטומט).

יש לנו את ה־δ של $L(M)$:

$$\delta(q_1,a_1)=q_{i_1},\delta(q_{i_1},a_2)=q_{i_2},\ldots,\delta\left(q_{i_{l-1}},a_l\right)\in F$$

אזי קיימות ההפקות המתאימות:

$$X_1\longrightarrow a_1X_{i_1},X_{i_1}\longrightarrow a_2X_{i_2},\ldots,X_{i_{l-1}}\longrightarrow a_l$$

לכן קיימת גזירה:

$$X_1\underset{\prod}{\Longrightarrow}a_1X_{i_1}\underset{\prod}{\Longrightarrow}a_1a_2X_{i_2}\overset{*}{\underset{\prod}{\Longrightarrow}}a_1a_2\ldots a_{l-1}X_{i_{l-1}}\underset{\prod}{\Longrightarrow}w$$

טענה 2 (כיוון ראשון): $w\in L(M)\Longleftarrow w\in L(\Gamma)$

בדיוק כמו מקודם, רק שהפעם, בכיוון השני: קיימת גזירה, לכן קיימות ההפקות המתאימות ויש לנו δ מתאימות להפקות ולכן המילה ב־ $L(M)$.

משפט: יהי Γ דקדוק רגולרי, אזי $L(\Gamma)$ רגולרית.

הוכחה: יהי $\Gamma = \{\sum, V, S, \prod\}$, כאשר: $V = \{S = X_1, \ldots, X_n\}$.

צ"ל כי קיים אסל"ד M כך ש: $L(\Gamma) = L(M)$ נגדיר את האסל"ד לפי החמישייה:

$\sum = \sum$, $Q = \{q_1, \ldots, q_n, q_F\}$, $F = \{q_F\}$,

נרצה למצוא את ה־δ המתאים עבור הפקה $X_i a \longrightarrow X_j \in \prod$. ה־$\delta$ המתאים יהיה: $X_i \longrightarrow aX_j \in \prod \Longrightarrow q_j \in \delta(q_i, a)$

עבור $X_i \longrightarrow a \in \prod$, ה־$\delta$ המתאים יהיה: $X_i \longrightarrow a \in \prod \Longrightarrow q_F \in \delta(q_i, a)$

טענה: $L(M) = L(\Gamma)$

טענה 1: $w \in L(M) \Longrightarrow w \in L(\Gamma)$

תהי מילה $w = a_1 \ldots a_l$, $w \in L(M)$ אזי $q_F \in \delta^*(q_1, w)$. אנו יודעים כי קיימים $q_{i_1}, q_{i_2}, \ldots, q_{i_l}$ כך ש:

$$q_1 \in \delta(q_{i_1}, a_1), q_{i_1} \in \delta(q_{i_2}, a_2), \ldots, q_F \in \delta(q_{i_{l-1}}, a_l)$$

אזי קיימות הפקות מתאימות:

$$X_1 \longrightarrow a_1 X_{i_1}, X_{i_1} \longrightarrow a_2 X_{i_2}, \ldots, X_{i_{l-1}} \longrightarrow a_l$$

לכן קיימת גזירה:

$$X_1 \underset{\prod}{\Longrightarrow} a_1 X_{i_1} \underset{\prod}{\Longrightarrow} a_1 a_2 X_{i_2} \overset{*}{\underset{\prod}{\Longrightarrow}} a_1 a_2 \ldots a_{l-1} X_{i_{l-1}} \underset{\prod}{\Longrightarrow} a_1 \ldots a_l = w$$

טענה 2: $w \in L(M) \Longleftarrow w \in L(\Gamma)$

בדיוק בכיוון השני: קיימת גזירה, לכן קיימות ההפקות המתאימות ויש לנו δ מתאימות להפקות ולכן המילה ב־$L(M)$.

25 למת הניפוח לשפות חסרות הקשר (גם: משפט בר־הילל)

יהי Γ דקדוק בצורת חומסקי בעל n משתנים. תהי z מילה ב־$L(\Gamma)$ כך שאורך z: $2^n < |z|$. מכאן, ניתן למצוא פירוק $z = uvwxy$ כך ש:

1. $0 < |vx|$
2. $|vwx| \leq 2^n$
3. לכל $0 \leq i$, מתקיים ש־ $uv^iwx^iy \in L(\Gamma)$

הוכחה: נגדיר עומק של עץ גזירה כמספר המשתנים המופיעים במסלוך הארוך ביותר בעץ. המילה הארוכה ביותר שניתן לגזור בעץ בעומק $n+1$ הוא 2^n. לכן, בעץ גזירה של המילה z, העומק יהיה לפחות $n+2$. לכן, במסלול הארוך ביותר בעץ קיים לפחות משתנה אחד X שמופיע לפחות פעמיים (לפי עקרון שובך יונים).

נעלה מהעלה העמוק ביותר עד שנתקל באותו משתנה פעמיים. נקרא לקודקוד שבו X מופיע בפעם הראשונה (התחתונה) α, ונקרא לקודקוד בו X מופיע בפעם השנייה (העליונה) β.

נסמן $\langle T_\alpha \rangle$ להיות w ($\langle T_\alpha \rangle$ = העלים של תת העץ שנגזר מהקודקוד α) , אזי: $X \underset{\prod}{\overset{*}{\Longrightarrow}} w$

$vwx = v \langle T_\alpha \rangle x = \langle T_\beta \rangle$, אזי: $X \underset{\prod}{\overset{*}{\Longrightarrow}} vXx$

$uvwxy = \langle T_s \rangle$ אזי: $S \underset{\prod}{\overset{*}{\Longrightarrow}} uXy \underset{\prod}{\overset{*}{\Longrightarrow}} uvXxy$,

ניתן להמשיך: $uv^iXx^iy \underset{\prod}{\overset{*}{\Longrightarrow}} \ldots \underset{\prod}{\overset{*}{\Longrightarrow}} uvvvXxxxy \underset{\prod}{\overset{*}{\Longrightarrow}} uvvXxxy \underset{\prod}{\overset{*}{\Longrightarrow}} uvXxy \underset{\prod}{\overset{*}{\Longrightarrow}}$ ולבסוף: $\underset{\prod}{\overset{*}{\Longrightarrow}}$

$uv^iXx^iy \underset{\prod}{\overset{*}{\Longrightarrow}} uv^iwx^iy$

העומק של $\langle T_\beta \rangle \le n+1$, לכן מס' העלים בו (לפי הטענה בהתחלה) $|\langle T_\beta \rangle| \le 2^n$, לכן $|vwx| \le 2^n$ (כי: $|\langle T_\beta \rangle| = |vwx|$)

משפט: תהי $L = \left\{ a^i b^i c^i \mid i > 0 \right\}$, אזי L אינה חסרת הקשר.

הוכחה: נניח בשלילה כי L חסרת הקשר. יהי Γ דקדוק בצורת חומסקי בעל n משתנים כך ש: $L(\Gamma) = L$. תהי $z = a^{2^n} b^{2^n} c^{2^n}$ אזי לפי משפט בר הילל, $z = uvwxy$ כך ש: $uwy \in L$ ו־2 התנאים $|vwx| \le 2^n$ ו־$|vx| > 0$ גם הם מתקיימים.

$$\overbrace{aa\ldots a}^{2^n}\overbrace{bb\ldots b}^{2^n}\overbrace{cc\ldots c}^{2^n}$$

אם ננסה לקחת "חלון" בגודל 2^n (לפי משפט בר הילל) אזי או שנמחק רק b, או שנמחק a ו־b או שנמחק b ו־c, ולכן לא נוכל לקחת כזה חלון, למחוק אותו ועדיין להשאר בשפה. סתירה.

26 תכונות הסגור של שפות חסרות־הקשר

26.1 סגירות לאיחוד

משפט: תהיינה L_1, L_2 שפות חסרות הקשר. אזי $L_1 \cup L_2$ היא חסרת הקשר.

הוכחה: יהי Γ_1 דקדוק ח"ה עם המרכיבים $\left\{ \sum, V_1, S_1, \prod_1 \right\}$ כך ש: $L(\Gamma_1) = L_1$ ויהי Γ_2 דקדוק ח"ה עם המרכיבים $\left\{ \sum, V_2, S_2, \prod_2 \right\}$ כך ש: $L(\Gamma_2) = L_2$ ונדאג לכך ש־ $V_1 \cap V_2 = \emptyset$ (נשנה את שמם אם צריך) .
צ"ל דקדוק ח"ה Γ כך ש: $L(\Gamma) = L_1 \cup L_2$

נגדיר את Γ : $\left\{ \sum, V, S, \prod \right\}$ כך ש:
$$\begin{gathered} V = V_1 \cup V_2 \cup S \\ \prod = \prod_1 \cup \prod_2 \cup \{ s \to s_1, s \to s_2 \} \end{gathered}$$

26.2 סגירות לחיתוך עם שפה רגולרית

משפט: תהי R שפה רגולרית. תהי L שפה ח"ה. אזי $L \cap R$ ח"ה.

הוכחה: יהי $\Gamma = \{\sum, V, S, \prod\}$ דקדוק ח"ה (על מנת להקל על עצמנו, ניקח דקדוק בצורת חומסקי) כך ש: $L(\Gamma) = L$ יהי $M = \{\sum, Q, q_1, F, \delta\}$ אס"ד כך ש: $L(M) = R$.

צ"ל דקדוק ח"ה $\Gamma' = \left\{\sum, V', S', \prod'\right\}$ כך ש: $L(\Gamma') = L \cap R$

נגדיר $V' = \{\theta^{pq} \mid \theta \in \sum \cup V, P \quad , \quad p, q \in Q\}$, S' חדש כלשהו, ו־ $\prod'$ בצורה הבאה:

1. עבור כל $q \in F$ מתקיים: $S' \longrightarrow S^{q_1 q}$ (אם ב־F יש לי 3 מצבים מקבלים אזי יש לי 3 הפקות מתאימות)
2. 2 תנאים:

   (א) לכל $X \longrightarrow YZ \in \prod$ ולכל $p, q, r \in Q$ מתקיים: $X^{pq} \longrightarrow Y^{pr} Z^{rq}$
   (ב) לכל $X \longrightarrow a \in \prod$ ולכל $p, q \in Q$ מתקיים: $X^{pq} \longrightarrow a^{pq}$
3. לכל $a \in \sum$ ולכל $p, q \in Q$ המקיימים $\delta(p, a) = q$ מתקיים: $a^{pq} \longrightarrow a$

נוכיח כי $L(\Gamma') = L \cap R$

תהי $w = a_1 \ldots a_l$ מילה כך ש: $w \in L \cap R$ (ונשים לב שמכיוון ש $w \in R$ אזי $\delta^*(q_1, w) = q_F$ כאשר q_F מצב מקבל כלשהו)

נבנה δ עבור R:
$$\begin{aligned} \delta(q_1, a_1) &= r_1 \\ \delta(r_1, a_2) &= r_2 \\ &\vdots \\ \delta(r_{l-1}, a_l) &= q_F \end{aligned}$$
. מכיוון ש־$w \in L$: $S' \underset{\prod'}{\Longrightarrow} S^{q_1 q_F} \underset{\prod'}{\Longrightarrow} \underbrace{a_1^{q_1 r_1} a_2^{r_1 r_2} a_3^{r_2 r_3} \ldots a_{l-1}^{r_{l-2} r_{l-1}} a_l^{r_{l-1} q_F}}_{r_1, r_2, \ldots, r_{l-1} \in Q}$

מכיוון ש־$w \in R$: $a_1^{q_1 r_1} a_2^{r_1 r_2} a_3^{r_2 r_3} \ldots a_{l-1}^{r_{l-2} r_{l-1}} a_l^{r_{l-1} q_F} \underset{\prod'}{\overset{*}{\Longrightarrow}} a_1 a_2 \ldots a_l = w$ והראינו מה שהיה להוכיח.

27 השפה L ח"ה אס"ם קיים אמל"ד M כך ש: $L(M) = L$

משפט: תהי L שפה ח"ה, אזי קיים אמל"ד M כך ש: $L(M) = L$

הוכחה: יהי Γ דקדוק חומסקי כך ש: $L(M) = L$. נבנה אמל"ד M כך ש: $L(M) = L(\Gamma)$

$\sum, Q = \{q_0, q_1\}, q_0, F = \{q_1\}, \prod = \{X_i \longrightarrow Y_i Z_i\} \cup \{x_j \longrightarrow q_j\}$

$$\begin{array}{c} q_0 \varepsilon \varepsilon S q_1 \\ g \in X_i, Y_i, Z_i, q_i \begin{cases} q_0 \varepsilon X_i Z_i p_i \\ p_i \varepsilon \varepsilon Y_i q_1 \end{cases} \\ q_1 a_j x_j \varepsilon q_1 \end{array}$$

דקדוק חומסקי:

$$\begin{array}{cc}
S \longrightarrow aSb & q_0\varepsilon\varepsilon Sq_1 \\
S \longrightarrow ab & q_1\varepsilon S\begin{cases}A\\B\end{cases} q_1 \\
S \longrightarrow AB & q_1\varepsilon S\begin{cases}A\\Z\end{cases} q_1 \\
A \longrightarrow a & q_1\varepsilon Z\begin{cases}S\\B\end{cases} q_1 \\
B \longrightarrow b & q_1aA\varepsilon q_1 \\
S \longrightarrow AZ & q_1bB\varepsilon q_1 \\
Z \longrightarrow SB & F=\{q_1\}
\end{array}$$

משפט: יהי $M=\{\sum,Q,q_0,F,\Omega,\delta\}$ אמל"ד , אזי $L\left(M\right)$ שפה ח"ה.

הוכחה:

הגדרה: אמל"ד M קרוי אמל"ד אטומי אם כל הפקודות ב־M הן מהצורות:

אוכל את $qa\varepsilon\varepsilon g'$

מוציא מהמחסנית $q\varepsilon j\varepsilon g'$

מכניס למחסנית $q\varepsilon\varepsilon jg'$

משפט: יהי M אמל"ד, אזי קיים אמל"ד אטומי M' כך ש: $L\left(M\right)=L\left(M'\right)$

הוכחה: תהי $qaSSg'$ פקודה כלשהי, נמיר את הפקודה בפקודות הבאות: $\begin{array}{c}qa\varepsilon\varepsilon p\\ p\varepsilon j\varepsilon r\\ r\varepsilon\varepsilon j'g'\end{array}$, וכך נוכל להמיר אמל"ד באמל"ד אטומי.

הגדרה: יהי M אמל"ד אטומי. <u>רישום</u> של חישוב על המילה w מוגדרת באופן איטרטיבי כך: אם הפקודה ה־i המבוצעת=

$$\begin{array}{ccc}
r_0=\varepsilon & & \\
r_1+1=r_i\cdot a_k & \Longleftrightarrow & qa_k\varepsilon\varepsilon g' \\
r_{i+1}=r_i\cdot)_k & \Longleftrightarrow & q\varepsilon J_k\varepsilon g' \\
r_{i+1}=r_i\cdot(_k & & q\varepsilon\varepsilon J_kg'
\end{array}$$

דוגמה: ניקח את המילה $abcba$:

$$\begin{array}{cc}
q_0a\varepsilon\varepsilon p_1 & \\
p_1\varepsilon\varepsilon Aq_0 & \\
q_0b\varepsilon\varepsilon p_2 & \\
p_2\varepsilon\varepsilon Bq_0 & \\
q_0c\varepsilon\varepsilon q_1 & a(_1b(_2b)_2a)_1 \\
q_1a\varepsilon\varepsilon p_3 & \\
p_3\varepsilon A\varepsilon q_1 & \\
q_1b\varepsilon\varepsilon p_4 & \\
p_4\varepsilon B\varepsilon q_1 &
\end{array}$$

טענה: חישוב מסתיים במחסנית ריקה $\Longleftrightarrow$ הסוגריים ברישום מאוזנים.

משפט: יהי M אמל"ד אזי $L(M)$ ח"ה.

הוכחה: הגדרנו רישום של חישוב (עבור אמל"ד אטומי).

טענה: חישוב מסתיים במחסנית ריקה אסם הסוגריים ברישום החישוב מאוזנים. (מקיים את חוקי הסוגריים: האחרון שפתחנו הוא הראשון שנסגור, וכל סוגריים שפתחנו ייסגרו בדיוק פעם אחת)

תהי $P_n = \{(_1,)_1,(_2,)_2,\ldots\}$ קבוצת סוגריים, ותהי $PAR_n \subset (\sum \cup P_n)^*$ כך ש:

$$PAR_n = \left\{w \in (\sum \cup P_n)^* \mid \text{parantessees in w are balanced}\right\}$$

טענה: PAR_n היא ח"ה.

הוכחה: צ"ל דקדוק ח"ה Γ כך ש- $L(\Gamma) = L$. (ומכאן $L = PAR_n$)

$$S \longrightarrow (_i S)_i \quad i = 1,\ldots,n$$
$$S \longrightarrow SS$$
$$S \longrightarrow a \qquad a \in \sum$$
$$S \longrightarrow \varepsilon$$

נשים לב שלא קיימת אפשרות לסוגריים בצורה בלתי חוקית, ולכן מ.ש.ל. על הטענה.

נמשיך בהוכחה: תהי $R = \left\{w \in (\sum \cup P_n)^* \mid \text{w is a register of a calculation, that ends with a recieving state}\right\}$

טענה: R היא רגולרית.

הוכחה: על מנת להוכיח את הטענה, נצטרך לבנות אסל"ד M' כך ש: $L(M') = R$

יהי M' כך ש: $M' = \left\{\sum', Q', q_0', F', \delta'\right\}$ כאשר $\begin{matrix}\sum' = \sum \\ Q' = Q \\ q_0' = q_0 \\ F' = F\end{matrix}$, נותר להגדיר δ: $\begin{matrix}\delta(q,a) = \{q' \mid qa\varepsilon\varepsilon q' \in M\} \\ \delta(q,(_i) = \{q' \mid q\varepsilon J_i \varepsilon q' \in M\} \\ \delta(q,)_i) = \{q' \mid q\varepsilon\varepsilon J_i q' \in M\}\end{matrix}$

מכאן R רגולרית.

נמשיך בהוכחה: ניעזר במשפט

תהי $\underbrace{ACC}_{\text{accepting}} = \left\{w \in (\sum \cup P_n)^* \mid \text{w is a recieving register}\right\}$

טענה: ACC היא ח"ה.

הוכחה: נשים לב ש:$ACC = R \cap PAR_n$. כעת, R רגולרית ו-PAR_n ח"ה, וחיתוך של שפה רגולרית עם שפה ח"ה היא ח"ה. מ.ש.ל. לטענה.

נמשיך בהוכחה: נגדיר $ERASE_{P_n}(w)$ כרישום המילה w ללא הסוגריים, ו-$ERASE_{P_n}(L)$ כשפה של אותן מילים.

נתבונן ב: $L(M) = ERASE_{P_n}(ACC)$ (נרצה להראות שאגף ימין יהיה ח"ה)

טענה: תהי L ח"ה. אזי $ERASE_{P_n}\left(L\right)$ ח"ה.

הוכחה: יהי Γ דקדוק חומסקי כך ש: $L\left(\Gamma\right)=L$

צריך למצוא דקדוק ח"ה Γ', כך ש: $L\left(\Gamma'\right)=ERASE_{P_n}\left(L\right)$

עבור כל $X\longrightarrow)_i$ או $X\longrightarrow (_i$ ב-Γ, נכניס ל-Γ' את $X\longrightarrow\varepsilon$. נשים לב שנקבל בדיוק את המילה, ללא הסוגריים.

נמשיך בהוכחה: אנו יודעים כי $L\left(M\right)=ERASE_{P_n}\left(ACC\right)$. אזי מכיוון ש-$ACC$ ח"ה , גם $L\left(M\right)$ ח"ה.