

מודלים חישוביים, חישוביות וסיבוכיות | 67521

הרצאות | פרופ' אורנה קופרמן

כתיבה | נמרוד רק

תשפ"ג סמסטר א'

תוכן העניינים

I	מבוא לאוטומטים	3
	הרצאה	3
	אוטומטים	4
	פעולות על שפות	8
	תרגול	8
II	אוטומטים אי-דטרמיניסטיים	13
	הרצאה	13
	תרגול	19
III	שפות לא רגולריות ולמת הניפוח	22
	הרצאה	22
	דוגמאות לשפות לא רגולריות	25
	תרגול	26
	ביטויים רגולריים	26
IV	משפט מיהיל-נרוד	29
	הרצאה	29
	מזעור אוטומטים	32
	תרגול	34
V	שפות חסרות הקשר	37
	הרצאה	37
	בעיית הריקנות ומשלים של אוטומט	38
	דקדוק חסר הקשר	39
	תרגול	43

שבוע II | מבוא לאוטומטים

הרצאה

חלק א' של ההרצאה

דוגמה נקפץ לחלק האחרון של הקורס (סיבוכיות). בהינתן גרף לא מכוון $G = \langle V, E \rangle$, נרצה לדעת האם יש בו מעגל אוילר (כזה שעובר בכל צלע בדיוק פעם אחת).

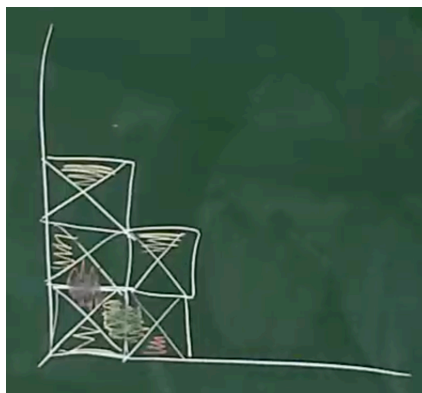
אוילר הוכיח שיש מעגל כזה אם ורק אם דרגת כל הקודקודים זוגית, ולכן ניתן להכריע את הבעיה בזמן לינארית כי יש לבעיה אפיון מתמטי. מעגל המילטון הוא מעגל שעובר בכל קודקוד בדיוק פעם אחת. לבעיה הזו אין אפיון מתמטי, והוכח שאין אלג' יותר טוב מאשר מעבר על כל האפשרויות, בסיבוכיות אקספוננציאלית.

דוגמה בהינתן $n = p \cdot q$, למצוא את p, q דורש זמן חישוב אקספוננציאלי באורך הייצוג, אפ'פ' שהאלג' הוא לינארי במספר עצמו. זה משום שהפרמטר שלנו במקרה הזה הוא לא המספר אלא הייצוג (אנחנו מקבלים $\log n$ ספרות/אחדים ואפסים, לא את המספר במלואו).

דוגמה קלט: $\{t_i\}$ אריכים שלכל אחד מהם יש צלעות $\{l_i\}, \{r_i\}, \{d_i\}, \{u_i\}$ (למעלה, למטה, ימינה ושמאלה בהתאמה) כאשר הצלעות הם צבעים (אדום, צהוב, ירוק).

פלט: האם ניתן לרצף באופן חוקי ריבוע $n \times n$ לכל $n \geq 1$, כאשר "חוקיות" מתבטאת בכך שצלעות סמוכות מסכימות על הצבע.

דוגמת ריצה באופן אינטואיטיבי, במקרים מסוימים, נוכל להציב אחד מהאריכים בפניה, למצוא אילו אריכים מתאימים לו מבחינת הצלעות הסמוכות, להציב אריכים חוקיים נוספים, וכך לחזור חלילה. לעתים (כמו זה שבתמונה), נוצרת תבנית של אריכים חוקיים על האלכסון (כלומר אריך א' בפניה השמאלית התחתונה, ואז ב' מימינו ומעליו, ואז ג' מימין ומעל כל ב') ואז אפשר לגדום את התבנית האינסופית הזו לריבוע $n \times n$ כל פעם שצריך ולהחזיר ריבוע חוקי. במקרה כזה הפלט יהיה כן.



איור 1: דוגמה לתבנית שנוצרת, אפשר להמשיך לצייר את האלכסון בכיוון דרום-מזרח ולחזור על התבנית החוצה עוד ועוד

הבעיה היא שאין שום ערובה לכך שהתבנית באמת קיימת במקרה הכללי, או שהיא נשמרת, ואי אפשר לרוץ עד ∞ . לכן התשובה היא שאין אלג' שפותר את הבעיה.

דוגמה (בעיית העצירה) קלט: תכנית מחשב P וקלט x .

פלט: האם P עוצרת על x .

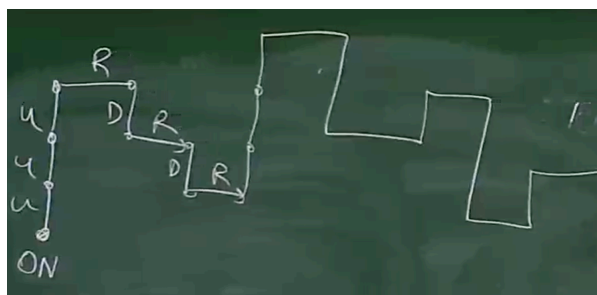
אין לבעיה זו אלג' שפותר אותה בכל המקרים (תחת הנחות מסוימות, אפשר לפעמים לתת תשובה).

אוטומטים

הגדרה אוטומט הוא מחשב עם זכרון מוגבל.

דוגמה נתון עט דיגיטלי שיכול לבצע אחת משש פקודות, ON, OFF, U, D, L, R . סדרת פקודות היא חוקית אם היא מתחילה ב- ON ,

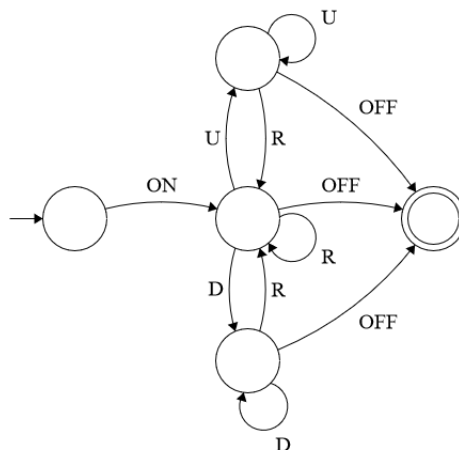
מסתיימת ב- OFF ומייצרת קו רקיע משמאל לימין.



איור 2: דוגמה לקו רקיע חוקי, אסור ללכת שמאלה ואסור לעלות מיד אחרי שיורדים (ולהפך)

נכתוב אוטומט שמחליט האם סדרת פקודות היא חוקית. אם נצליח לעבור בין המצבים (העיגולים), החל מהמצב הראשון (זה עם חץ

ללא מקור) ועד למצב המקבל (עם העיגול הכפול) על קשתות קיימות, הרי שהסדרה חוקית.



איור 3: אוטומט חוקי

אינטואיטיבית, המצב האמצעי הוא זה שממנו אפשר לעשות מה שרוצים, העליון הוא אחרי עלייה והתחתון הוא אחרי ירידה. נשים

לב כי מכולם אפשר לפנות ימינה.

הגדרה אוטומט (automaton, DFA) הוא חמישייה $A = \langle Q, \Sigma, \delta, q_0, F \rangle$ שהם המצבים, הא"ב, פונקציית המעברים, המצב ההתחלתי וקבוצת המצבים המקבלים שמוכלת ב- Q .

• δ היא פ' $Q \times \Sigma \mapsto Q$.

• Σ היא קבוצה סופית של אותיות, לדוגמה $\{0, 1\}^4$, $\Sigma = \{0, 1\}$ וכו'.

• מילה היא $w = w_1, \dots, w_n$ סדרה סופית של אותיות, ו- ϵ היא המילה הריקה.

• שפה היא קבוצה של מילים, $L \subseteq \Sigma^*$, כאשר w מילה סופית מעל הא"ב Σ : $\Sigma^* = \{w : w \text{ מילה סופית מעל הא"ב } \Sigma\}$.

דוגמה A_1 הוא האוטומט בצירור. במקרה הזה $\Sigma = \{0, 1\}, Q = \{q_0, q_1\}, F = \{q_0\}$ ופ' המעברים היא.

δ	0	1
q_0	q_0	q_1
q_1	q_1	q_0

הגדרה ריצה על מילה $w = w_1 \dots w_n$ מעל Σ היא סדרה של מצבים $r = r_0 \dots r_n$ כך ש:

• $r_0 = q_0$ (הריצה מתחילה ב- q_0).

• לכל $i \geq 0$, $r_{i+1} = \delta(r_i, w_{i+1})$ (הריצה מכבדת את δ).

דוגמה עבור A_1 והמילה 011, הריצה היא $q_0 q_1 q_0$.

הגדרה r היא ריצה מקבלת (accepting) אם $r_n \in F$ (המצב האחרון בריצה הוא מקבל). אחרת, r הוא דוחה (rejecting).

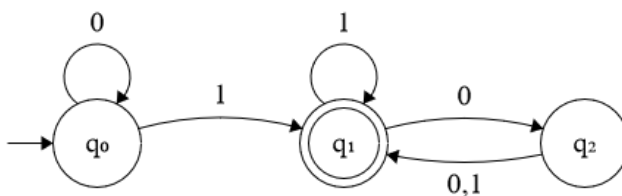
A מקבל את w אם הריצה של A על w היא מקבלת.

$L(A)$, השפה של האוטומט היא אוסף המילים ש- A מקבל עליהן.

דוגמה עבור A_1 , $L(A_1) = \{w : w \text{ הוא זוגי}\}$ (אפשר להוכיח באינדוקציה).

הערה אם לא קיים מעבר עבור אות ומצב, אפשר או להחליט ש- δ לא מוגדרת על כל $Q \times \Sigma$ או להחליט שכל קשת לא קיימת מובילה לבור דוחה, כלומר מצב לא מקבל שאי אפשר לצאת ממנו.

דוגמה נצייר אוטומט נוסף, A_2 , ונחשב את השפה שלו.



איור 4: האוטומט A_2

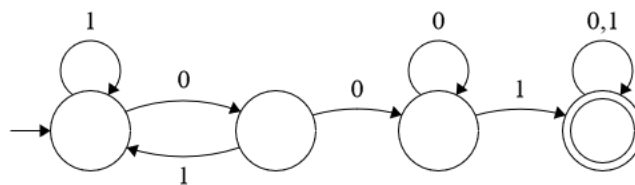
נסמן בצבע האם כמה מילים נבחרות הן בשפה או לא, 010, 011, 001110, 1, 11, 00000.

אם נחשוב עוד קצת, נגלה ש-

$$L(A_2) = \{w : \text{יש ב- } w \text{ לפחות 1 אחד, ואחרי ה-1 האחרון יש מספר זוגי (או אפס) של 0-ים}\}$$

בתרגול נוכיח את זה באופן פורמלי.

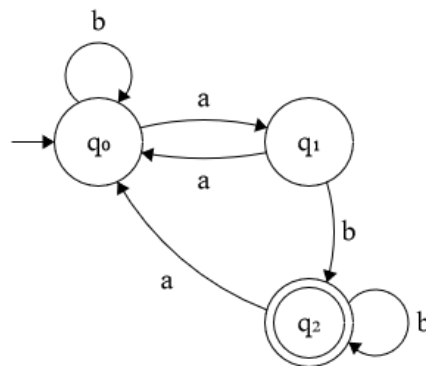
דוגמה בהינתן שפה, ננסה לחשב את האוטומט. השפה היא $\{w : 001 \text{ מכילה את הרצף } w\}$.



איור 5: אוטומט שנגזר מ- L_3

חלק ב' של ההרצאה

דוגמה $L = \{w : \#_a w \wedge w_n = b\}$, $\Sigma = \{a, b\}$ (כאשר $\#_a w$ הוא מספר ה- a -ים ב- w ו- w_n האות האחרונה במילה).



איור 6: אוטומט שאנחנו טוענים שנגזר מ- L

המצב ההתחלתי לא מקבל כי $b \notin L$. כאות ראשונה לא מקדם אותנו כי זו לא מילה חוקית. הרעיון בהלוך-חזור ב- q_0, q_1 הוא שרק אם המספר הוא אי זוגי של a -ים, נגיע ל- q_1 ומשם נעצור במצב מקבל רק אם אנחנו נגמרים ב- b .

לכל מצב נוכל להתאים סטטוס - מה מאפיין את המילה שמגיעה אליו (לאחר מכן נשתמש בסטטוסטים האלה, נפרמל אותם ונוכיח איתם את נכונות האוטומט):

• $q_0 - \#_a w$ זוגי.

• $q_1 - \#_a w$ אי זוגי ו- w מסתיימת ב- a .

• $q_2 - \#_a w$ אי זוגי ו- w מסתיימת ב- b .

טענה $L(A) = L$.

הוכחה: $\forall w \in \Sigma^*$ (אוסף המילים האפשריות) מתקיים $\delta^*(q_0, w) = q_0$ (כאשר $\delta^* : Q \times \Sigma^* \mapsto Q$), כלומר הפעלה שוב ושוב של δ על המילה החל ממצב נתון).

נוכיח את שלוש הטענות הבאות ומשם ינבע כי

$$w \in L \iff \delta^*(q_0, w) = q_2 \iff \delta^*(q_0, w) \in F$$

האם נובע משתי הטענות הראשונות, הטענה השלישית מספקת לנו רק כיוון אחד.

1. אם $\delta^*(q_0, w) = q_0$ אז $\#_a w$ זוגי.

2. אם $\delta^*(q_0, w) = q_1$ אז $\#_a w$ אי זוגי ו- w מסתיימת ב- a .

3. אם $\delta^*(q_0, w) = q_2$ אז $\#_a w$ אי זוגי ו- w מסתיימת ב- b .

באינדוקציה על $|w|$:

בסיס ($|w| = 0$): $w = \epsilon$ ולכן $\delta^*(q_0, \epsilon) = q_0$ ואכן $\#_a \epsilon$ זוגי.

צעד ($|w| \rightarrow |w| + 1$): נוכיח את הטענה על $w \cdot a$, $w \cdot b$ בהנחה שהיא נכונה על w . נוכיח רק את המקרה של $w \cdot a$ ונשאיר לסטודנטית המשקיעה להוכיח את המקרה השני.

• אם $\delta^*(q_0, w \cdot a) = q_0$ אז בהכרח $\delta^*(q_0, w) \in \{q_1, q_2\}$ ולכן מה"א $\#_a w$ אי זוגי (מטענות 2 ו-3) ולכן $\#_a w \cdot a$ זוגי.

– אם $\delta^*(q_0, w \cdot a) = q_1$ אז $\delta^*(q_0, w) = q_0$ ומה"א $\#_a w$ זוגי ולכן $\#_a w \cdot a$ אי זוגי ו- w נגמרת ב- a .

– אם $\delta^*(q_0, w \cdot a) = q_2$ אז זה לא ייתכן (מהגדרת האוטומט).

■

דוגמה $L = \{w : \#_a w = \#_b w\}$, $\Sigma = \{a, b\}$. אין אוטומט סופי ששפתו L ! זה בגלל שאחרי ה- a הראשון, נצטרך "לזכור" שיש לנו 1 לטובת a , ואז אם שוב יש a נצטרך לזכור עוד 1, ואם b אז אחד לטובת b וזה אינסופי בעצם.

הגדרה שפה רגולרית היא שפה שניתנת לזיהוי ע"י אוטומט, ונסמן $L \in \text{REG}$, פורמלית, L היא רגולרית אם קיים DFA כך ש- $L(A) = L$.

פעולות על שפות

תהינה $L_1, L_2 \in \Sigma^*$. כל הפעולות עובדות על שפות מעל $\Sigma_1 \neq \Sigma_2$ ובמקרה כזה נסמן $\Sigma = \Sigma_1 \cup \Sigma_2$.

1. איחוד (union): $L_1 \cup L_2 = \{w : w \in L_1 \vee w \in L_2\}$ (שפות הן קבוצות, זו לא פעולה חדשה).

2. שרשר (concatenation): $L_1 \cdot L_2 = \{w_1 \cdot w_2 : w_1 \in L_1, w_2 \in L_2\}$ (הצמדה של כל צמד מילים משתי השפות).

3. כוכב (star): $L^* = \{w_1 \cdot \dots \cdot w_k : k \geq 0 \wedge w_i \in L, \forall i \leq k\}$ (שרשר של 0 או יותר מילים ב- L כולל את ϵ עבור $k = 0$).

דוגמה $L_1 = \{1, 333\}, L_2 = \{22, 4444\}$

$$L_1 \cup L_2 = \{1, 333, 22, 4444\}$$

$$L_1 \cdot L_2 = \{122, 1444, 33322, 3334444\}$$

$$L^* = \{\epsilon, 1, 333, 11, 1333, 3331, 333333, \dots\}$$

הערה אם $L = \emptyset$ אז $L^* = \{\epsilon\}$ וכך גם עבור $L = \{\epsilon\}$. כל שפה אחרת היא אינסופית (יש לפחות מילה אחת לא ריקה, נשרר אותה כמה פעמים שרק נרצה).

תרגול

הגדרה נאמר כי $R \subseteq S \times T$ הוא יחס מעל S, T (לרוב $S = T$).

דוגמה $R = \{(a, b) : |a - b| \leq 1\}, A = \{1, 2, 3, 4\}$

תכונות של יחסים

• רפלקסיביות: $(a, a) \in A, \forall a \in A$ או בסימון חלופי, aRa (היחס הנ"ל הוא רפלקסיבי).

• סימטריה: aRb או bRa (היחס הנ"ל הוא סימטרי).

• טרנזיטיביות: aRb ו- bRc אז aRc .

• יחס שקילות: יחס שמקיים את שלושת הנ"ל.

יחס שקילות R מעל A מחלק את A למחלקות שקילות זרות המוגדרות ע"י $[a]_R = \{b \in A : aRb\}$, כי אם קיים $x \in [a]_R \cap [b]_R$ אבל $[a]_R \neq [b]_R$ אז קיים $c \in [a]_R \setminus [b]_R$ ולכן

$$aRc \Rightarrow cRx \Rightarrow cRb \Rightarrow c \in [b]_R$$

סתירה.

דוגמה $G = \langle V, E \rangle$ גרף לא מכוון והיחס $R \subseteq V \times V$ שמשמעותו "כל זוגות הקודקודים שיש ביניהם מסלול ב- G ". קל לראות שזהו יחס רפלקסיבי, טרנזיטיבי וסימטרי ולכן זהו יחס שקילות.

הגדרה עוצמה של קבוצה היא מדד ל"גודל" הקבוצה. עבור קבוצה סופית A , העוצמה שלה היא $|A|$.

הגדרה $|\mathbb{N}| = \aleph_0$.

הערה ראינו ש- $|\mathbb{Z}| = |\mathbb{Q}| = \aleph_0$ (כאשר שוויון עוצמות משמעו קיום פ' חזע"ל בין שתי הקבוצות).

הערה נאמר כי $|A| \leq |B|$ אם יש העתקה חח"ע מ- A ל- B ו- $|A| < |B|$ אם בנוסף אין העתקה חח"ע מ- A על B .

טענה (האלכסון של קנטור) $\aleph_0 < 2^{\aleph_0} = |[0, 1]|$.

הגדרה $\Sigma^n = \underbrace{\Sigma \times \dots \times \Sigma}_n$ ונגדיר $\Sigma^* = \bigcup_{n=0}^{\infty} \Sigma^n$

הערה רבים מתבלבלים כאן אבל חשוב לזכור ש- Σ סופית וכך גם Σ^n , אבל Σ^* אין סופית.

דוגמאות לשפות

$$\Sigma = \{a, b\}$$

$$\bullet L_1 = \{\epsilon, a, aa, b\}$$

$$\bullet L_2 = \{w : w_1 = a\} \text{ (מילים שמתחילות ב-} a\text{)}$$

$$\bullet L_3 = \{\epsilon\}$$

$$\bullet L_4 = \emptyset \text{ וזו אינה אותה קבוצה כמו } L_3!$$

$$\bullet L_5 = \{w : |w| < 24\}$$

• $L_1 = \{w : w_1 = a\}$ ו- $L_2 = \{w : w_n = b\}$ (סימון לקוני למילים שמסתיימות ב- b). שפה נוספת היא

$$L_1 \cup L_2 = \{w : w_1 = a \vee w_n = b\}$$

$$L_2 \cdot L_1 = \{w : ab \text{ מכילה את הרצף } w\}$$

$$L_1 \cap L_2 = \{w : w_1 = a \wedge w_n = b\}$$

$$L_1 \cdot L_2 = L_1 \cap L_2$$

כאשר השוויון האחרון נכון כי המילה הראשונה בצמד מתחילה ב- a והשנייה נגמרת ב- b ובאמצע לא משנה מה יש, בדומה ל- $L_1 \cap L_2$.

$$L = \{ww : w \in \Sigma^*\} \quad \bullet$$

$$\bar{L} = \Sigma^* \setminus L = \{w : 2 \nmid |w|\} \cup \{w = w_1 \dots w_{2n} : w_1, \dots, w_n \neq w_{n+1} \dots w_{2n}\}$$

$$L \cdot L = \{w w x x : w, x \in \Sigma^*\}$$

הערה כל שפה מקיימת $L \subseteq \Sigma^*$, או באופן שקול $L \in P(\Sigma^*)$.

כמה מילים יש ב- Σ^* ? $|\Sigma^*| = \aleph_0$.

כמה שפות יש מעל Σ^* ? $2^{|\Sigma^*|} = 2^{\aleph_0}$.

כמה שפות רגולריות יש מעל Σ^* ? \aleph_0 , כי כל אוטומט מוגדר ע"י מחרוזות מעל א"ב סופי (המצבים, הא"ב של האוטומט וכו') ולכן

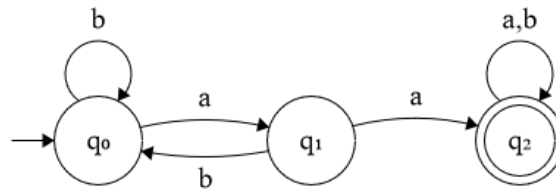
מהנ"ל עוצמת אוסף המחרוזות ששקולות לאוטומטים היא \aleph_0 . לחלופין, כל אוטומט אפשר לצייר ויש מספר בן מנייה של פיקסלים

על canvas (במחשב).

מסקנה קיימות שפות לא רגולריות, ויש "יותר" לא רגולריות מאשר לא (השפות הרגולריות הן קבוצה במידה 0 מתוך כל השפות).

$$\delta^*(q, w) = \begin{cases} q & w = \epsilon \\ \delta(\delta^*(q, w'), \sigma) & w = w'\sigma, \sigma \in \Sigma \end{cases} \quad \text{הגדרה בהינתן אוטומט } A, \text{ נגדיר}$$

דוגמה נביט באוטומט הבא.



איור 7: אוטומט לדוגמה

נחשב ערך של δ^* .

$$\delta^*(q_1, ba) = \delta(\delta^*(q, b), a) = \delta(\delta(\delta^*(q, \epsilon), b), a) = q_1$$

דוגמה עבור $\Sigma = \{0, \dots, 9, \#\}$ והשפה

$$L = \{x\#a : x \in \{0, \dots, 0\}^*, a \in \{0, \dots, 9\}, a \in x\}$$

נמצא את האוטומט המתאים ל- L . ראשית נשים לב לדוגמה כי $64424\#5 \notin L$ אבל $1243\#2 \in L$.

הבעיה באוטומט זה שאין לנו זיכרון ולכן נצטרך "לזכור" מספיק מידע כדי לזכור האם ראינו $\#$ עד עכשיו ואילו ספרות ראינו עד כה.

נבחר $Q = (2^{\{0, \dots, 9\}} \times \{1, 2\}) \cup \{q_{acc}, q_{sink}\}$ כאשר מצב מייצג את אוסף הספרות שראינו עד כה והאם ראינו את סולמית עד עכשיו (2 כן ראינו).

$q_0 = \langle \emptyset, 1 \rangle$ כלומר לא ראינו את סולמית ולא ראינו אף ספרה, $F = \{q_{acc}\}$ ו- $\Sigma = \{0, \dots, 9, \#\}$.

$$\delta(\langle c, i \rangle, \sigma) = \begin{cases} \langle c \cup \{\sigma\}, 1 \rangle & \sigma \in \{0, \dots, 9\}, i = 1 \\ \langle c, 2 \rangle & \sigma = \#, i = 1 \\ q_{acc} & \sigma \in c, i = 2 \\ q_{sink} & \sigma \notin c, i = 2 \end{cases}$$

עברו על כל המצבים והבינו את המשמעות, הרעיון בסוף הוא שאם ראינו סולמית ונתקלנו באות נוספת, נקבל או נשלול בהתאם להאם ראינו את הספרה או לא. לשם השלמות גם נגדיר $\delta(q_{acc}, \sigma) = \delta(q_{sink}, \sigma) = q_{sink}$ כי אם הגענו למצב המקבל והוספנו עוד תו זה כבר לא בשפה.

טענת עזר בהינתן $w \in \{0, \dots, 9\}^*$, נגדיר $S(w) = \{\sigma \in \{0, \dots, 9\}^* : w \text{ מופיעה ב-}\sigma\}$ (אוסף הספרות שמופיעות ב- w). נוכיח כי $\delta^*(q_0, w) = \langle S(w), 1 \rangle$.

הוכחה: באינדוקציה על $|w|$.

בסיס ($w = \epsilon$): $\delta^*(q_0, w) = \delta(q_0, \epsilon) = \langle \emptyset, 1 \rangle = \langle S(w), 1 \rangle$ כנדרש.

צעד ($|w| - 1 \rightarrow |w|$): נסמן $w' = w\sigma$

$$\delta^*(q_0, w') = \delta(\delta^*(q_0, w), \sigma) \stackrel{\text{נ"ח}}{=} \delta(\langle S(w), 1 \rangle, \sigma) = \langle S(w'), 1 \rangle$$

■

טענה $L = L(A)$.

הוכחה: נוכיח הכלה דו-כיוונית באינדוקציה על אורך המילה; זו דרך ההוכחה המקובלת לטענות על שפות ואוטומטים.

$L \subseteq L(A)$: נניח כי $w \in L$ ונראה שריצה של A על w מקבלת. כאשר $w = x\#a$, $x \in [0, \dots, 9]^*$, $a \in \{0, \dots, 9\}$ ו- $a \in S(x)$.

$$\begin{aligned}\delta^*(q_0, w) &= \delta(\delta^*(q_0, x\#), a) \\ &= \delta\left(\delta\left(\frac{\delta^*(q_0, x), \#}{\langle S(x), 1 \rangle}, a\right)\right)\end{aligned}$$

$$\delta \text{ הגדרת } = \delta(\langle S(x), 2 \rangle, a)$$

$$\delta \text{ הגדרת } = q_{acc}$$

$L(A) \subseteq L$: מספיק שנוכיח שאם $w \notin L(A)$ אז $w \notin L$. נעבור על כל המילים $w \notin L(A)$.

• אם $w \in \{0, \dots, 9\}^*$ אז מטבעת העזר $\delta(q_0, w) = \langle S(w), 1 \rangle \neq q_{acc}$.

• אם $w \in \{0, \dots, 9\}^* \times \{\#\}$ אז

$$\delta^*(q_0, w) = \delta\left(\frac{\delta^*(q_0, w), \#}{\langle S(x), 1 \rangle}\right) = \langle S(x), 2 \rangle \neq q_{acc}$$

• אם $w = x\#y$ עבור $|y| > 1$ אז

$$\delta^*(q_0, w) = \delta^*(\langle S(x), 2 \rangle, y) \neq q_{acc}$$

כאשר השוויון נובע מכך שניתן לפצל את הריצה על $x\#$ ואז על y . הריצה על $x\#$ מביאה אותנו ל- $\langle S(x), 2 \rangle$ מהגדרה של δ . האי-שוויון נובע מכך ש- $|y| > 1$ ולכן גם אם אחרי הספרה הראשונה של y הגענו ל- q_{acc} , בהכרח כל הספרות האחרות יובילו אותנו תמיד לבור דוחה.

• אם $w = x\#a$ אבל $a \notin S(x)$ אז

$$\begin{aligned}\delta^*(q_0, w) &= \delta(\delta(\delta^*(q_0, x), \#), a) \\ &= \delta(\langle S(x), 2 \rangle, a)\end{aligned}$$

$$a \notin S(x) \Rightarrow q_{sink}$$

שבוע III | אוטומטים אי-דטרמיניסטיים

הרצאה

חלק א' של ההרצאה

משפט השפות הרולגריות סגורות לאיחוד, כלומר, אם $L_1, L_2 \in \text{REG}$ אז $L_1 \cup L_2 \in \text{REG}$.

הוכחה: בהינתן DFA-ים $A_1 = \langle Q_1, \Sigma, \delta_1, s_1, F_1 \rangle$, $A_2 = \langle Q_2, \Sigma, \delta_2, s_2, F_2 \rangle$, נבנה $A = \langle Q, \Sigma, \delta, s_0, F \rangle$ שעבורו $L(A) = L(A_1) \cup L(A_2)$.

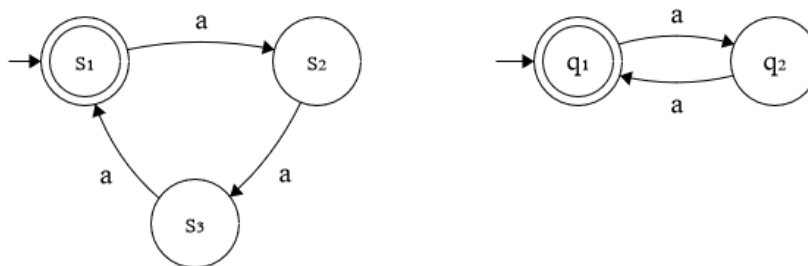
הרעיון הוא ש- A מסמלץ את A_1 ו- A_2 יחד, ואוטומט בבנייה כזו נקרא אוטומט המכפלה. נבחר $Q = Q_1 \times Q_2$, $s_0 = \langle s_1, s_2 \rangle$, ופ' מעברים

$$\delta(\langle q_1, q_2 \rangle, \sigma) = \langle \delta_1(q_1, \sigma), \delta_2(q_2, \sigma) \rangle$$

כאשר אנחנו מניחים ש- A_1, A_2 לא נתקעים כי אפשר להוסיף בור דוחה במקרה הצורך.

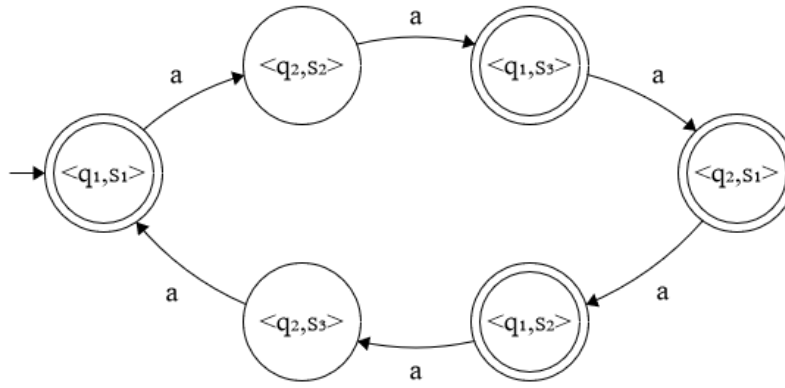
הערה אם $L \subseteq \{a\}^*$ אז היא מגדירה תת קבוצה של \mathbb{N} - כל האורכים של מילים בשפה, כלומר $\{i : a^i \in L\}$.

דוגמה נבחר את האוטומטים A_1, A_2 כבתמונה,



איור 8: האוטומטים A_1 (מימין) ו- A_2 (משמאל)

במקרה הזה, אוטומט המכפלה יראה כבאיור, כאשר בכל מעבר אנחנו "צועדים" קדימה גם במצבים של A_1 וגם בשל A_2 .



איור 9: A אוטומט המכפלה

ולא קשה לראות שהאוטומט מקבל על מספרים זוגיים וכאלה שמתחלקים בשלוש, כלומר $L(A) = \{w : |w| \bmod 2 = 0 \vee |w| \bmod 3 = 0\}$.

הערה מהדוגמה הנ"ל ניתן לראות שאם היינו רוצים לבנות אוטומט שהשפה שלו היא $L(A_1) \cap L(A_2)$ היינו בוחרים

$$F = \{\langle q_1, q_2 \rangle : q_1 \in F_1 \wedge q_2 \in F_2\}$$

כאשר ההבדל כאן הוא "וגם" במקום "או" על המצבים המקבלים.

הערה אם היינו רוצים A עם $L(A) = \Sigma^* \setminus L(A_1)$, מספיק שהיינו מגדירים $A = \langle Q_1, \Sigma, \delta_1, s_1, Q_1 \setminus F_1 \rangle$ כי הריצה מגיעה ל- $Q_1 \setminus F_1$ אם "אם" A_1 דוחה את w .

נוכיח כי $L(A) = L(A_1) \cup L(A_2)$. תהי $w = w_1 w_2 \dots w_n$ מילה ב- Σ^* ותהי $r = r_0 r_1 \dots r_n$ הריצה של A על w . נסמן $r_i = \langle q_1^i, q_2^i \rangle$. מהגדרת A , $q_1^0 = s_1, q_2^0 = s_2$ ולכן $i \geq 0$,

$$q_1^{i+1} = \delta_1(q_1^i, w_i), q_2^{i+1} = \delta_2(q_2^i, w_i)$$

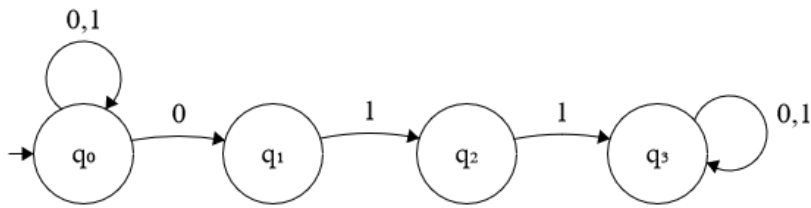
ולכן $\rho_1 = q_1^0, q_1^1, \dots, q_1^n$ היא ריצה של A_1 על w ובהתאמה $\rho_2 = q_2^0, q_2^1, \dots, q_2^n$ היא ריצה של A_2 על w .

מכאן, r מקבלת אם "אם" $\langle q_1^n, q_2^n \rangle \in F$ או $q_1^n \in F_1$ או $q_2^n \in F_2$ אם "אם" r_1 מקבלת או r_2 מקבלת אם "אם" $w \in L(A_1)$ או $w \in L(A_2)$.

■

הערה בדרך להוכחה ש-REG סגור לשרשור, נתקעים בקושי הוכחתי. לכאורה נפרק מילה לשני החלקים, נריץ כל חלק באוטומט המתאים לו ונסיים. הבעיה היא שלכל מילה יכולים להיות כמה פירוקים. לשם כך נצטרך "לנחש" מתי לקפוץ.

אוטומטים אי-דטרמיניסטיים



איור 10: אוטומט אי-דטרמיניסטי

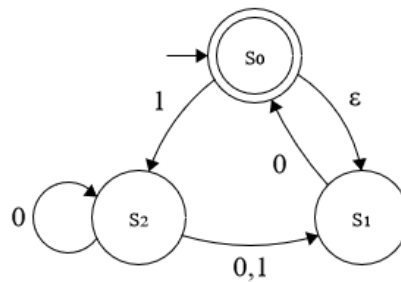
לכאורה פ' המעברים לא מוגדרת היטב עבור $q_0, 0$, אבל כאן הרעיון הוא שהאוטומט יכול לבחור מתוך כמה אפשרויות בעצמו לאיזה מצב הוא עובר, כאשר מילה מתקבלת ע"י האוטומט אם קיימת ריצה עם ניחשים כלשהם שמקבלת, ובמקרה כזה נגדיר $\delta(q_0, 0) = \{q_0, q_1\}$.

הגדרה אוטומט אי-דטרמיניסטי הוא אוטומט שבו פ' המעברים ממפה מצב ואות (או אפסילון) לקבוצה של מצבים עוקבים אפשריים, כלומר

$$\delta : Q \times (\Sigma \cup \{\epsilon\}) \rightarrow 2^Q$$

ומילה מתקבלת אם קיימת ריצה מקבלת של A על המילה.

דוגמה נביט באוטומט הבא עם "צעד אפסילון",



איור 11: אוטומט אי-דטרמיניסטי עם "צעד אפסילון"

המילים הבאות מתקבלות: $\epsilon, 0, 00, 00110$ (כי נוכל להשתמש קודם בצד אפסילון במקום ליפול לבור דוחה מ- s_0) ואילו $001, 00111$ לא מתקבלות.

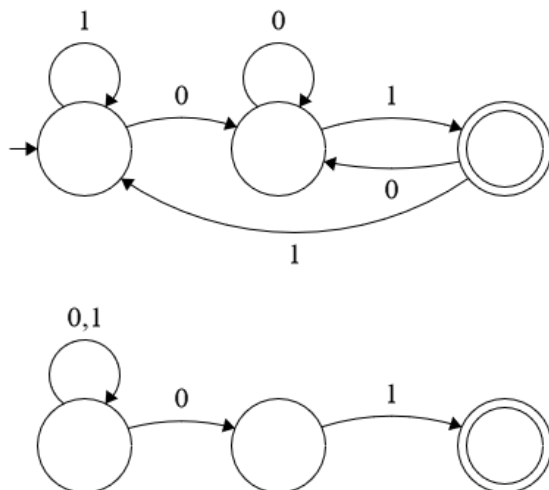
הגדרה אוטומט אי-דטרמיניסטי הוא חמשייה מהצורה $A = \langle Q, \Sigma, \delta, Q_0, F \rangle$ שעבורה $Q_0 \subseteq Q$ (יכולים להיות כמה מצבים התחלתיים)

$$\delta : Q \times (\Sigma \cup \{\epsilon\}) \rightarrow 2^Q$$

ריצה של A על מילה $w = \sigma_1 \sigma_2 \dots \sigma_n$ היא סדרת מצבים $r = r_0 r_1 \dots r_m$ (כאשר $m \geq n$ בגלל ריפודי אפסילון) כך שניתן לכתוב את w כ- $w' = x_1 x_2 \dots x_m$ כאשר $x_i \in \Sigma \cup \{\epsilon\}$ ומתקיים $r_0 \in Q_0$ וכן $r_{i+1} \in \delta(r_i, x_{i+1})$ (בניגוד ל- $=$ ב-DFA). בנוסף, $r_m \in F$ מקבלת אם.

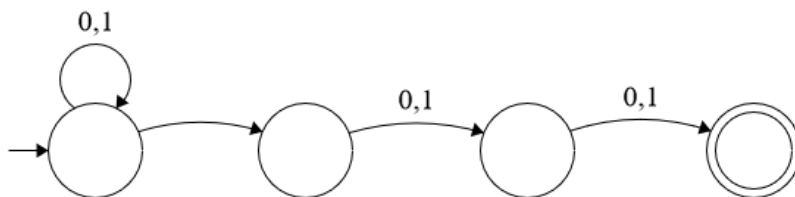
נאמר כי A מקבלת את w אם קיימת ריצה של A על w שמקבלת.

דוגמה NFA מעל $\Sigma = \{0, 1\}$. $L = \{w : w \text{ מסתיימת ב- } 0, 1\}$, באיור למעלה DFA שהשפה שלו היא L ולמטה NFA שקול (ויותר פשוט),



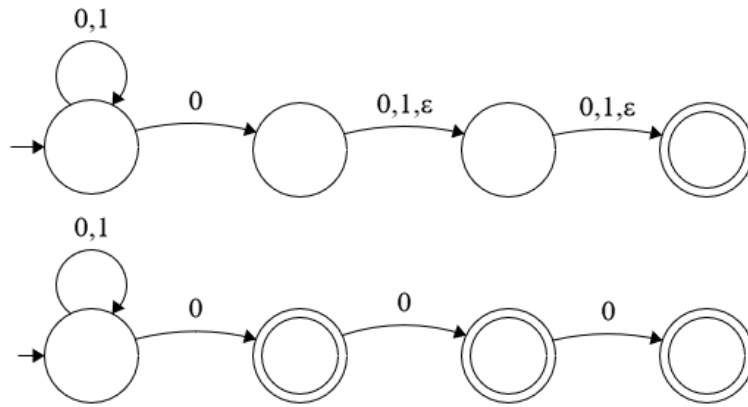
איור 12: אוטומט דטרמיניסטי (למעלה) ואי-דטרמיניסטי (למטה) שמשרתים אותה המטרה

דוגמה עבור w מסתיימת ב- $L = \{w : 0(0+1)(0+1)\}$, האוטומט הבא מקבל אם מילה היא ב- L (הוכחה פורמלית פשוטה בעל פה),



איור 13: אוטומט עם השפה הנ"ל

דוגמה עבור $\{0\}$ במקום הלפני לפני אחרון, הלפני אחרון או האחרון, $L' = \{w : w \text{ האוטומטים הבאים הם בעלי השפה } L'\}$,



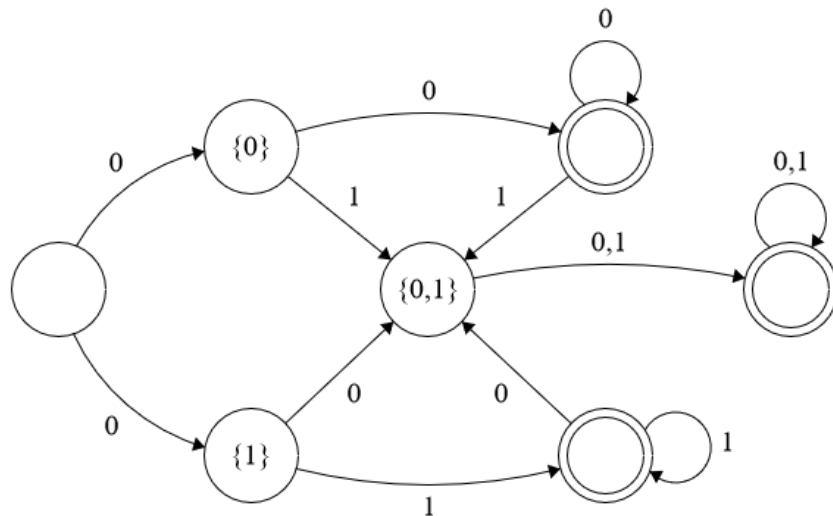
איור 14: שני אוטומטים אי-דטרמיניסטיים ששפתם L'

דוגמה מצבים התחלתיים רבים הם שימושיים לדוגמה במקרה של אוטומט המכפלה, שם אם היינו יכולים להגדיר כמה מצבים התחלתיים יכולנו לעשות בניה יותר פשוטה עם $Q = Q_1 \cup Q_2$.

ההוכחה למשפט בסוף ההרצה עבר לתחילת חלק ב' של ההרצה.

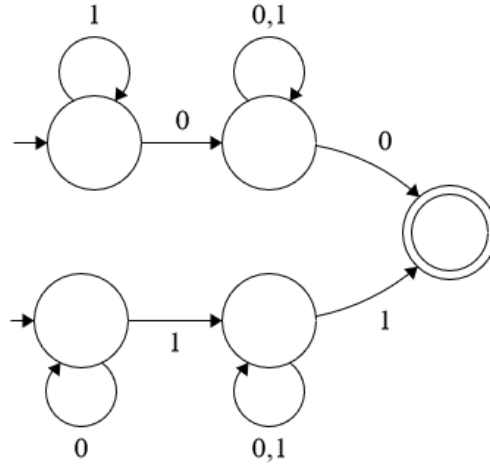
חלק ב' של ההרצה

דוגמה L היא השפה שבה כל המילים שבהן האות האחרונה הופיע לפניכן במילה, מעל $\Sigma = \{0, 1\}$. ראו DFA שמתאים לה באיור,



איור 15: DFA שמתאים ל- L

ועתה NFA מתאים (שקול), כאשר הרעיון כאן הוא שהחלק העליון מתאים לריצה שבה יש 0 אחד לפחות ובסוף 0 ולמטה זו כזו בהתאם שמסתיימת ב-1.



איור 16 : NFA שמתאים ל- L

משפט לכל NFA A קיים DFA A' שקול כך ש- $L(A) = L(A')$.

הוכחה: בהינתן $A = \langle Q, \Sigma, Q_0, \delta, F \rangle$, נבנה $A' = \langle Q', \Sigma, q'_0, \rho, F' \rangle$ כך ש- $L(A) = L(A')$. נבחר $Q' = 2^Q$ ואז הרעיון הוא ש- A' מגיע למצב S בריצה אחרי קריאת w אם ורק אם A יכול להגיע לבדיקת כל המצבים ב- S אחרי קריאת w .

באופן אינדוקטיבי, δ^* מוגדרת ע"י $\delta^*(s, \epsilon) = s$, $\delta^*(s, \sigma) = \bigcup_{s' \in S} \delta(s', \sigma)$, $\delta^*(S, \sigma) = \bigcup_{s \in S} \delta^*(s, \sigma)$, ובצעד ה- n י, $\delta^*(S, w \cdot \sigma) = \delta^*(\delta^*(S, w), \sigma)$.

נבחר $q'_0 = Q_0$ שהוא קבוצה, אבל $q'_0 \in Q'$ כי Q' זו קבוצה של קבוצות ולכן זה בסדר.

נגדיר $\rho(S, \sigma) = \bigcup_{s \in S} \delta(s, \sigma)$ לכל $s \in Q'$ ו- $\sigma \in \Sigma^*$.

טענה לכל $w \in \Sigma^*$ מתקיים $\rho^*(q'_0, w) = \delta^*(Q_0, w)$ או במילים, המצב ב- A' ש- A' מגיע אליו אחרי קריאת w (המצב הוא קבוצה בפני עצמו), שווה לקבוצת המצבים ש- A יכול להיות בה (באחת הריצות שלו) על A .

נבחר $F' = \{S \in 2^Q : S \cap F \neq \emptyset\}$ כי אנחנו מקבלים אם הגענו למצב ב- Q' שאחד מ(תתי-)המצבים שבו הם מקבלים (כי זה אומר שאנחנו יכולים להגיע אליו בריצה כלשהי של A').

נוכיח כי $L(A) = L(A')$. $w \in L(A)$ אם ורק אם קיימת ריצה מקבלת של A על w אם ורק אם $\delta^*(Q_0, w) \cap F \neq \emptyset$ (עכשיו נוכיח) $w \in L(A)$ אם ורק אם $\rho^*(q'_0, w) \in F'$.

הוכחה: (של הטענה המקוננת) באינדוקציה על w :

בסיס $(w = \epsilon)$: $\rho^*(q'_0, \epsilon) = q'_0 = Q_0 = \delta^*(Q_0, \epsilon)$.

צעד $(|w| \rightarrow |w| + 1)$:

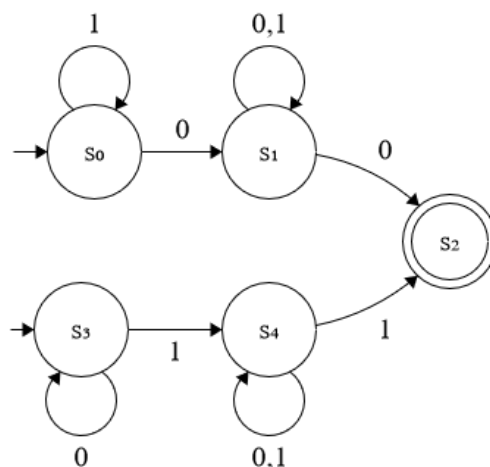
$$\rho^*(q'_0, w \cdot \sigma) = \rho(\rho^*(q'_0, w), \sigma) \stackrel{\text{הגדרת } \delta^*}{=} \delta^*(\rho^*(q'_0, w)) \stackrel{\text{ה"ח}}{=} \delta^*(\delta^*(Q_0, w), \sigma) = \delta^*(Q_0, w \cdot \sigma)$$

■

■

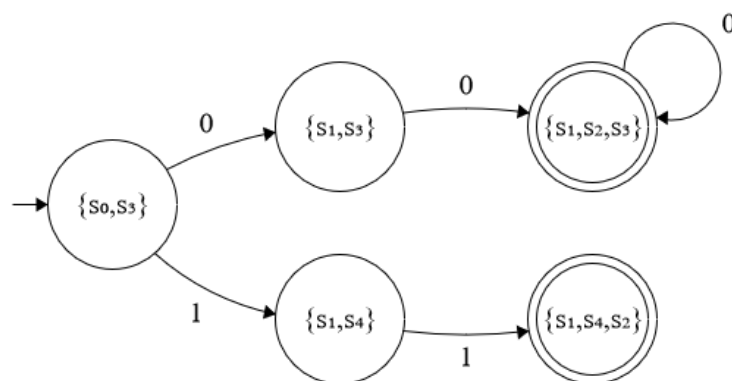
וזה מסיים את ההוכחה כי השפות של האוטומטים שוות.

דוגמה בחזרה לדוגמה למעלה (מצורף איור נוסף), נמצא DFA מתאים לזה (נבצע דטרמיניזציה).



איור 17 : NFA שראינו למעלה

ה-DFA המתאים הוא כבאיור, כאשר הוא לא שלם כי יש 2^5 מצבים. הרעיון בכל אופן הוא שבכל פעם אנחנו מסתכלים לאן כל אחד מהמצבים לוקח אותנו בהינתן האות הנוכחית ואוספים את כולם לכדי מצב (כמו ההגדרה של ρ), ושמצב הוא מקבל אם "ס" הוא מכיל מצב שהיה מקבל ב-NFA.



איור 18 : DFA חלקי שמתאים ל-NFA למעלה

תרגול

טענה לכל NFA $A = \langle Q, \Sigma, \delta, Q_0, F \rangle$ קיים NFA שקול ב- B כך שב- B אין מעבר ϵ .

הוכחה: הרעיון הוא שנקבץ את כל המצבים שעוברים אליהם עם ϵ לאחד כל פעם ונראה שזה שקול. נגדיר

$$E(q) = \{s : \epsilon \text{ עם מעברי } s \text{ ב-} A \text{ רק עם מעברי } \epsilon\}$$

נשים לב כי תמיד $q \in E(q) \neq \emptyset$ ובפרט $E(q) \neq \emptyset$ (לא לצעוד מ- q זה כמו לצעוד אפסילון מ- q כי לא קראנו כלום).

נגדיר $B = \left\langle Q, \Sigma, \delta', \bigcup_{q \in Q_0} E(q), F \right\rangle$ כאשר הרעיון במצבים ההתחלתיים הוא כל המצבים שאפשר להגיע אליהם ממצב התחלתי כלשהו רק בצעדי אפסילון.

נגדיר $\delta'(q, \sigma) = \bigcup_{s \in \delta(q, \sigma)} E(s)$ כלומר כל מצב שאפשר להגיע אליו עם האות ומעברי אפסילון מ- q (בפרט זה כולל גם את מצבי $\delta(q, \sigma)$ המקוריים).

לא נוכיח נכונות אבל כן נסביר למה הבניה הזו היא פולינומיאלית: אפשר לחשב כל $E(q)$ בזמן יעיל באמצעות DFS כאשר קשת קיימת בגרף שלנו אם-היא מעבר אפסילון בין שני מצבים באוטומט. ■

טענה REG סגורה לאיחוד, כלומר אם $L_1, L_2 \in \text{REG}$ אז $L_1 \cup L_2 \in \text{REG}$.

הוכחה: יהיו $A_1 = \langle Q_1, \Sigma, \delta_1, q_1, F_1 \rangle, A_2 = \langle Q_2, \Sigma, \delta_2, q_2, F_2 \rangle$ DFA-ים ל- L_1, L_2 בהתאמה. בה"כ $Q_1 \cap Q_2 = \emptyset$ (אפשר לשנות את השמות, זה לא מעניין). נבנה NFA B לשפה $L_1 \cup L_2$ כך ש-

$$B = \langle Q_1 \cup Q_2, \Sigma, \delta, \{q_1, q_2\}, F_1 \cup F_2 \rangle$$

ופ' המעברים מוגדרת ע"י

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

כך, מילים מ- L_1 יוכלו להתקבל מריצות שמתחילות ב- q_1 ומילים ב- L_2 מתקבלות על ריצות החל מ- q_2 (למעשה יש לנו שני אוטומטים זרים שכל ריצה יכולה לבחור איפה היא מתחילה).

נראה ש- $L(B) = L_1 \cup L_2$ באמצעות הכלה דו כיוונית.

$\underline{L_1 \cup L_2 \subseteq L(B)}$: תהי $w \in L_1 \cup L_2$ ובה"כ $w \in L_1$. היות ש- $w \in A_1$ על w מקבלת ונסמנה $r_0, \dots, r_{|w|}$ כאשר $r_0 = q_1$ ו- $r_{|w|} \in F_1$. נשים לב שהריצה $r_0, \dots, r_{|w|}$ היא ריצה אפשרית של B על w כי $r_0 \in \{q_1, q_2\}$, פ' המעברים δ מוגדרת היטב במקרה ש- $q \in Q_1$ וזה מתקיים לכל אורך המסלול ובנוסף $r_{|w|} \in F_1 \subseteq F_1 \cup F_2$ ולכן הריצה גם מקבלת, כלומר $w \in L(B)$.

$\underline{L(B) \subseteq L_1 \cup L_2}$: תהי $w \in L(B)$, כלומר קיימת ריצה מקבלת של B על w שנסמנה $r_0, \dots, r_{|w|}$. מהגדרת B , $r_0 \in \{q_1, q_2\}$ ונני בה"כ כי $r_0 = q_1$. היות ש- $q_0 \in Q_1$, לכל $0 \leq i \leq |w| - 1$ מתקיים שהמעבר מ- r_i ל- r_{i+1} נעשה דרך הפ' δ_1 מוגדרת רק על מצבים מ- Q_1 ו- $Q_1 \cap Q_2 = \emptyset$ ולכן התמונה של δ_1 מוכל ב- Q_1 . בנוסף, $r_{|w|} \in F_1 \cup F_2$ אבל $r_{|w|} \in Q_1$ ולכן $r_{|w|} \in F_1$, לכן, $r_0, \dots, r_{|w|}$ היא גם ריצה מקבלת של A_1 על w ולכן $w \in L_1$. ■

טענה REG סגורה לשרשור, כלומר אם $L_1, L_2 \in \text{REG}$ אז $L_1 L_2 \in \text{REG}$.

הוכחה: הרעיון הוא שנאפשר קפיצה (בניחוש) מכל מצב מקבל ב- A_1 להתחלה של A_2 ואז כך נאפשר שרשור של מילים.

יהיו $A_1 = \langle Q_1, \Sigma, \delta_1, q_1, F_1 \rangle, A_2 = \langle Q_2, \Sigma, \delta_2, q_2, F_2 \rangle$ DFA-ים ל- L_1, L_2 בהתאמה. בה"כ $Q_1 \cap Q_2 = \emptyset$. נגדיר NFA B לשפה $L_1 \cdot L_2$ ע"י $B = \langle Q_1 \cup Q_2, \Sigma, \delta, \{q_1\}, F_2 \rangle$ כאשר

$$\delta(q, \sigma) = \begin{cases} \delta_1(q, \sigma) & q \in Q_1, \sigma \in \Sigma \\ \delta_2(q, \sigma) & q \in Q_2, \sigma \in \Sigma \\ \{q_0\} & q \in F_1, \sigma = \epsilon \end{cases}$$

נוכיח הכלה דו כיוונית. **הוכחה:** $\underline{L_1 \cdot L_2 \subseteq L(B)}$: תהי $w \in L_1 \cdot L_2$ כלומר $w = x \cdot y$ כאשר $x \in L_1, y \in L_2$. ישנן ריצות מקבלות של A_1 על x ושל A_2 על y , בהתאמה נסמנן $r_0, \dots, r_{|x|}$ ו- $r'_0, \dots, r'_{|y|}$. נשים לב כי הריצה $r'_0, \dots, r'_{|y|}$ היא ריצה של B על y . $w = x \cdot y$ ■

$r_0 = q_0$ הוא אכן מצב התחלתי ב- B ועד $r_{|x|}$ הריצה של B על x ממשיכה כמו זו של A_1 ומסתיימת ב- F_1 ו- $r_{|x|} \in F_1$.

מכאן יש מעבר $\delta(r_{|x|}, \epsilon) = \{q_0\}$ ומשם הריצה של B על ההמשך של w שהוא בדיוק y היא כמו של A_2 על y .

זו מסתיימת ב- F_2 ו- $r_{|y|} \in F_2$ ובגלל שהמצבים המקבלים של B הם גם F_2 , קיבלנו ריצה מקבלת.

$\underline{L(B) \subseteq L_1 \cup L_2}$: תהי $w \in L(B)$ ותהי $r_0, \dots, r_{|w|}$ ריצה מקבלת (כלשהי) של B על w . מתקיים $r_0 = \{q_1\}$ ו- $r_{|w|} \in F_2$. מהגדרת B , כדי להגיע ל- F_2 חייב להיות קיים $k \in [|w|]$ כך שהמעבר $r_k \rightarrow r_{k+1}$ השתמש במעבר ϵ ממצב ב- F_1 ל- $\{q_2\}$.

נביט במילים $x = w_1, \dots, w_k$ ו- $y = w_{k+1}, \dots, w_{|w|}$. מהגדרת B , הריצה r_0, \dots, r_k היא ריצה של A_1 על x שמסתיימת ב- F_1 ולכן זו ריצה מקבלת של A_1 על x ולכן $x \in L(A_1)$.

באופן דומה, הריצה של B על y החל מ- r_{k+1} היא ריצה של A_2 על y שמסתיימת במצב מקבל ב- F_2 ולכן $y \in L(A_2)$ ולכן

$$w \in L(A_1) \cdot L(A_2)$$

■

טענה REG סגורה לפעולה Kleene-Star כלומר אם $L \in \text{REG}$ אז $L^* = \bigcup_{k \in \mathbb{N}_0} \underbrace{L \cdot \dots \cdot L}_k \text{ פעמים } k$ $L^* \in \text{REG}$

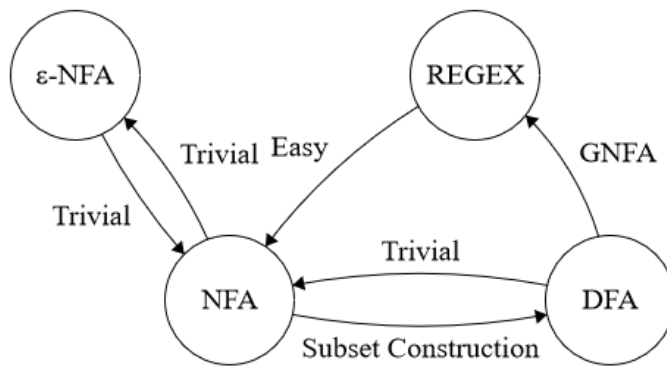
הוכחה: יהי $A = \langle Q, \Sigma, \delta, q_0, F \rangle$ DFA ל- L . לכאורה היינו יכולים לבנות NFA שהוא A פשוט עם חיבור מהמצבים הסופיים למצב ההתחלתי שוב עם צעד אפסילון. הבעיה היא שאם A לא מקבל את המילה הריקה, גם הבניה לא אבל $\epsilon \in L^*$. לכן נוסף מצב נוסף q_{start} שהוא יהיה המצב ההתחלתי היחיד שיש ממנו צעד אפסילון למצב ההתחלתי של A .

נבנה NFA B לשפה L^* ע"י $B = \langle Q \cup \{q_{start}\}, \Sigma, \delta', \{q_{start}\}, \{q_{start}\} \rangle$ כאשר בה"כ $q_{start} \notin Q$. δ מוגדרת ע"י

$$\delta'(q, \sigma) = \begin{cases} \{\delta(q, \sigma)\} & q \in Q \\ \emptyset & q = q_{start} \\ \hline \emptyset & q = q_{start} \wedge \sigma = \epsilon \\ \{q_{start}\} & q \in F \wedge \sigma = \epsilon \\ \{q_0\} & q = q_{start} \wedge \sigma = \epsilon \end{cases}$$

■

הערה ראו איור של מצבנו מבחינת שקילות של אוטומטים, כאשר בקרוב נלמד על REGEX-ים,



איור 19: מפת שקילות בין אוטומטים

שבוע IIII | שפות לא רגולריות ולמת הניפוח

הרצאה

חלק א' של ההרצאה

הערה בהרצאה הקודמת הראנו איך לעשות דטרמיניזציה ל-NFA וראינו שבהינתן A' NFA עם n מצבים, ל-DFA השקול לו יש לכל היותר 2^n מצבים (חסם עליון).

היום נראה שאין פולינום p שבהינתן p (כל) NFA עם n מצבים, יש לו DFA שקול עם לכל היותר $p(n)$ מצבים (חסם תחתון). מקרים פרטיים כמו כן יכולים להיות חסומים ע"י פולינום בגדילה שלהם כשהם DFA, אבל שום בנייה לא תעבוד לכל NFA אפשרי.

הערה לא מספיק שנראה, לדוגמה, שפה L כך שיש ל- L NFA עם 10 מצבים, אבל כל DFA עבור L צריך 2^{10} מצבים.

זה לא מוכיח שום דבר כי זה לא סותר את הפולינום $p(n) = n^3 + 500$, שעבורו $p(10) > 2^{10}$ ולכאורה הוא מצליח לחסום NFA-ים כלשהם (כמו כן שלא את כולם).

משפט לכל פולנום p , קיימת שפה L כך של- L קיים NFA עם n מצבים וה-DFA הקטן ביותר עבור L צריך יותר $p(n)$ מצבים.

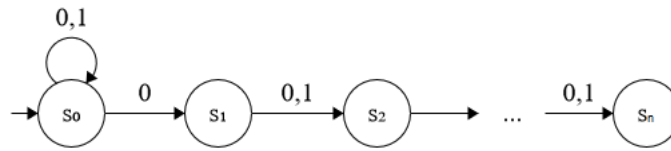
הוכחה: מספיק שנראה שלכל $n \geq 1$ קיימת L_n כך של- L_n קיים NFA עם $n + 1$ מצבים אבל ה-DFA הקטן ביותר עבור L_n צריך לפחות 2^n מצבים.

כי אם בשלילה קיים פולינום כאמור, נתבונן ב- n_0 שמובטח שעבורו $2^{n_0} > p(n_0 + 1)$. משם ה-DFA הקטן ביותר עבור L_{n_0} מכיל $2^{n_0} > p(n_0 + 1)$ מצבים כפי שנוכיח עכשיו בסתירה לקיום פולינום שמקיים את התנאים.

נבחר $\Sigma = \{0, 1\}$ ונגדיר

$$L_n = (0 + 1)^* 0 (0 + 1)^{n-1} = \{w : 0 \text{ היא מהסוף ה-} n \text{ האות ה-} n\}$$

כאשר הביטוי משמאל נקרא ביטוי רגולרי - $0, 1$ כמה פעמים שנרצה (רישא), 0 , ואז $n - 1$ או 0 -ים. ראו איור של NFA מתאים לשפה,



איור 20: NFA ל- L_n

נניח בשלילה שיש DFA D_n כך ש- $L(D_n) = L_n$ ויש ל- D_n פחות מ- 2^n מצבים. ישנם 2^n וקטורים באורך n מעל $\{0, 1\}$ ולכן 2^n מילים שונות באורך n מעל הא"ב $\{0, 1\}$.

אם ב- D_n יש פחות מ- 2^n מצבים, אז מעקרון שובך היונים יש שתי מילים $w_1 \neq w_2 \in (0 + 1)^n$ שעבורן D_n מגיע לאותו המצב בסוף קריאתן. ופורמלית, עבור $D_n = \langle \{0, 1\}, Q, q_0, \delta, F \rangle$, קיימות $w_1 \neq w_2 \in (0 + 1)^n$ כך ש-

$$q = \delta^*(q_0, w_1) = \delta^*(q_0, w_2)$$

מהיות $w_1 \neq w_2$, הרי שקיים $i \in [n]$ כך ש- $w_1[i] \neq w_2[i]$ ובה"כ $w_1[i] = 0, w_2[i] = 1$. נוכיח שבהכרח האוטומט טועה כי נשרשר סיפא למילים כך ש- i יהיה האינדקס ה- n מהסוף ואז האוטומט מסווג את שתי המילים באותה הדרך בניגוד לכך שאחת הוא אמור לקבל והאחרת לדחות (מהגדרת השפה). נתבונן ב- $s = \delta^*(q, 1^{i-1})$.

• אם $s \in F$ אז D_n מקבל את $w_2 \cdot 1^{i-1}$ בסתירה לנכונות D_n , שכן $w_2 \cdot 1^{i-1} \notin L$ (האות ה- n מהסוף היא $w_2[i-1]$).

• אם $s \notin F$ אז D_n דוחה את $w_1 \cdot 1^{i-1}$ (הוא DFA והריצה היחידה מגיעה ל- s) בסתירה לנכונות D_n , שכן $w_1 \cdot 1^{i-1} \in L$.

■

כלומר הגענו לסתירה בכל המקרים.

טענה אין DFA עבור $L = \{0^n 1^n : n \geq 0\}$.

הוכחה: נניח בשלילה כי $A = \langle Q, \Sigma, q_0, \delta, F \rangle$ הוא DFA עם $L(A) = L$. יהי $p = |Q|$. נתבונן במילה $w = 0^p 1^p \in L$ ולכן הריצה של A על w , $r = q_0 q_1, \dots, q_{2p}$, מקבלת, כלומר $q_{2p} \in F$.

ברישא q_0, \dots, q_p יש מעגל, כלומר קיימים $0 \leq l < j \leq p$ כך ש- $q_l = q_j$ (מעקרון שובך היונים). לכן יש ל- A ריצה מקבלת גם על $0^{p-(j-l)} 1^p \notin L$ (כי אפשר לגדום את המעגל מ- l ל- j ולהסתכל על הריצה $q_0 \dots q_l q_{j+1} \dots q_{2p}$). ■

משפט (למת הניפוח לשפות רגולריות, pumping lemma) אם L רגולרית אז קיים $p \geq 1$ (קבוע הנפוח) כך שלכל מילה $w \in L$, אם $|w| \geq p$ אז קיימת חלוקה $w = x \cdot y \cdot z$ כך ש:

$$1. |x \cdot y| \leq p.$$

$$2. |y| > 0 \text{ (} y \neq \epsilon \text{)}.$$

$$3. \forall i \geq 0, xy^i z \in L, \text{ המילה } L.$$

הערה אם L סופית אז אפשר לקחת $p = l + 1$ עבור l אורך המילה הארוכה ביותר ב- L ואז הלמה מתקיימת באופן ריק.

דוגמה עבור $L = (0 + 1)^* 0 (0 + 1)$ (כל המילים שהאות הלפני אחרונה שלהם היא 0). ניקח $p = 3$ ונתבונן במילה $w \in L$ עם $|w| \geq 3$. נבחר $w = x \cdot y \cdot z$ כאשר $|x| = \epsilon, |y| = 1, |z| = |w| - 1$.

אכן $y \neq \epsilon$ וכמוכן $|x \cdot y| = 1 \leq 3$ ולכן $xy^i z \in L, i \geq 0$ כי $|z| \geq 2$ ולכן האות הלפני האחרונה ב- $xy^i z$ נשארת האות הלפני האחרונה ב- z , הלא היא 0.

הוכחה: תהי L שפה רגולרית. יהי A DFA שמזהה את L ונבחר p להיות מספר המצבים ב- A . נתבונן במילה $w \in L$ עם $|w| \geq p$. בריצה של A על w , יש מצב שחוזר בקריאת p האותיות הראשונות, כלומר קיימים $0 \leq j < l \leq p$ כך ש- $q_l = q_j$ (מעקרון שובך היונים). נבחר x, y, z "ע"

$$w = \frac{w_1 \dots w_j}{x} \frac{w_{j+1} \dots w_l}{y} \frac{w_{l+1} \dots w_n}{z}$$

ונראה שהתנאים של הלמה מתקיימים. $j < l$ ולכן $|y| > 0$ וגם $|x \cdot y| \leq p$ כי $l \leq p$ ואכן $xy^i z \in L$ כי הריצה היא

$$q_0, \dots, q_j, (q_{j+1} \dots q_l)^i, q_{l+1}, \dots, q_n$$

כאשר זו ריצה חוקית כי יש מעבר מ- q_l ל- q_{l+1} . ■

חלק ב' של ההרצאה

הערה נוכל להשתמש בשלילת למת הניפוח כדי להוכיח ששפות הן לא רגולריות. אם למת הניפוח מספרת לנו ש- $\alpha \in \text{REG}$, אז $\neg \alpha$ הוא שלכל $p \geq 1$, קיימת מילה $w \in L$ עם $|w| \geq p$ כך שלכל חלוקה $w = x \cdot y \cdot z$, אם $|x \cdot y| \leq p$ וגם $|y| > 0$, קיים $i \geq 0$ כך ש- $xy^i z \notin L$.

או במילים, לכל קבוע ניפוח קיימת מילה ארוכה מהקבוע כך שלא משנה איזו חלוקה נבחר עם $y \neq \epsilon$ ו- $|xy| \leq p$, אחד הניפוחים של y לא בשפה.

את הבחירה על השלילה של שלושת התנאים עשינו כי זה נוח אבל אפשר היה גם לעשות שאם 1, 3 מתקיימים אז 2 לא מתקיים.

דוגמאות לשפות לא רגולריות

1. תהי $L_1 = \{0^n 1^n : n \geq 0\}$. זו שפה לא רגולרית (ראינו כבר אבל גם) כי לכל p , נוכל להתבונן במילה $0^p 1^p$. לכל חלוקה $xyz = 0^p 1^p$

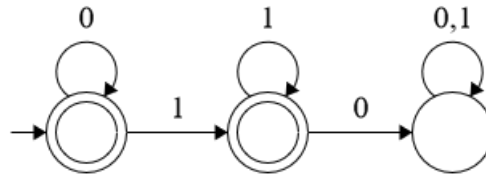
כך ש- $|xy| \leq p$, מתקיים $y = 0^j$ עבור $1 \leq j$ (אחרת xy זולג ל-1-ים ויוצא שהוא ארוך מ- p). לכן, $xy^2 z = 0^{p+j} 1^p \notin L_1$.

2. $L_2 = \{w : \#_0 w = \#_1 w\}$ היא גם לא רגולרית. ההוכחה הנ"ל עובדת גם כן כי גם שם $0^p 1^p \in L$ ו- $|0^p 1^p| \geq p$ וכו' וכו'.

יש דרכים אחרות בהינתן שידוע לנו ש- L_1 לא רגולרית להוכיח ש- L_2 לא רגולרית.

• ניסיון 1: $L_1 \subseteq L_2$ ו- L_1 לא רגולרית ולכן L_2 לא רגולרית - לא עובד! $(0+1)^*$ אבל האחרונה כן רגולרית (DFA טריוויאלי).

• ניסיון 2: עבור $L_3 = 0^* 1^*$ קיים DFA שמזהה אותה (ראו איור). מתקיים $L_1 = L_2 \cap L_3$ ומסגירות שפות רגולריות לחיתוך, נובע ש- L_2 לא רגולרית (אחרת החיתוך שלה עם L_3 היה רגולרי בסתירה לכך ש- L_1 לא רגולרית).



איור 21: DFA ל- L_3

3. $L_4 = \{0^n 1^m : n > m\}$ לא רגולרית לפחות אינטואיטיבית. נוכיח זאת עם למת הניפוח. בהינתן p , נתבונן במילה $0^{p+1} 1^p$ ובחלוקה

xyz כך ש- $|xy| \leq p$, $|y| > 0$, בהכרח $y = 0^j$ עבור $j \geq 1$. הניפוח עם $i = 0$ מוציא מהשפה (ניפוח מטה), כי

$$0^{p+1-j} 1^p = xy^0 z = xz$$

אבל $p+1-j \leq p$ וזה לא בשפה.

4. $L_5 = \{w \cdot w : w \in (0+1)^*\}$ היא לא רגולרית (אינטואיטיבית) ונראה זאת עם למת הניפוח. בהינתן p , נתבונן במילה

$w = 0^p 10^p 1 \in L_5$ ואכן $|w| \geq p$. לכל חלוקה $w = xyz$ כך ש- $|xy| \leq p$ ו- $|y| > 0$ מתקיים $y = 0^j$ עבור $j \geq 1$ (כרגיל) ונתבונן

ב- $i = 2$, שעבורו $xy^2 z = 0^{p+j} 10^p 1$ היא מילה לא בשפה (הצדדים שלה לא שווים).

5. $L_6 = \{a^p : p \text{ ראשוני}\}$ (כאשר $\Sigma = \{a\}$) היא לא רגולרית (כלומר גם אין אפיון עם מספר מצבים סופי של המספרים הראשוניים).

בהינתן p , יהי q ראשוני עם $q > p$. נתבונן במילה $w = a^q$ ותהי חלוקה $w = xyz$ כך ש- $|xy| \leq p$ ו- $|y| > 0$. נסמן $|x| = n$, $|y| = m$ ולכן

$$|z| = q - (n + m)$$

$$|xy^i z| = n + mi + q - (n + m) = m(i - 1) + q$$

ועבור $i = q + 1$ מתקיים $|xy^iz| = m((q + 1) - 1) + q = (m + 1)q$ וכן $m > 0$ ולכן $(m + 1) > 1$.

6. $\Sigma = \{0, 1\}$ עבור $L_7 = \{w : \text{פלינדורם } w\}$. נתבונן ב- $0^p 10^p$ ואז אם $|xy| \leq p$ ו- $|y| > 0$ אז $xy^2z \notin L$ כי ה-1 לא באמצע.

תרגול

ביטויים רגולריים

הגדרה ביטוי רגולרי מעל א"ב Σ הוא אחד מהבאים :

• \emptyset .

• ϵ .

• $a \in \Sigma$.

• t, s כאשר $t^*, t \cup s, t \cdot s$ ביטויים רגולריים קצרים יותר.

הערה דרך נוספת לייצג ביטוי רגולרי מעל $\{a, b\}$ היא $r := \emptyset | \epsilon | a | b | r \cup s | r \cdot s | r^*$.

דוגמה נביט בביטוי מעל $\Sigma = \{a, b\}$, $(a \cup b)^* bb (a \cup b)^*$. השפה שלו היא כל המילים שמכילות את הרצף bb .

דוגמה הביטוי $(1^* \cup 2^*)^* 00^*$ מייצג את כל המילים שמתחילות באחד או יותר אפסים ונגמרות ברצף כלשהו של 1-ים או של 2-ים.

הגדרה בהינתן ביטויים רגולריים r, s, t , נגדיר את השפה שלהם כך :

• אם $r = \emptyset$ אז $L(r) = \emptyset$.

• אם $r = \epsilon$ אז $L(r) = \{\epsilon\}$.

• אם $r = a \in \Sigma$ אז $L(r) = \{a\}$.

• אם $r = s \cdot t$ אז $L(r) = L(s) \cdot L(t)$.

• אם $r = s \cup t$ אז $L(r) = L(s) \cup L(t)$.

טענה $L \in \text{REG}$ אם קיים ביטוי רגולרי r כך ש- $L = L(r)$.

הוכחה: \Rightarrow יהי r ביטוי רגולרי ונראה שקיים NFA A_r כך ש- $L(r) = L(A_r)$ באינדוקציה על אורך סדרת היצירה של r (מספר התווים בכתובה של הביטוי הרגולרי, כך ϵ הוא באורך 1 לדוגמה).

• אם $r = \emptyset$ אז נבחר A_r להיות NFA ריקה (ששפתו ריקה).

• אם $r = \epsilon$ אז ניקח את A_r להיות NFA ששפתו היא $\{\epsilon\}$ (לדוגמה אוטומט שהמצב ההתחלתי שלו הוא מקבל וכל אות מובילה לבור דוחה).

- אם $r = a \in \Sigma$ אז נבחר A_r להיות NFA ששפתו היא $\{a\}$ (מצב התחלתי לא מקבל, מעבר ממנו למצב מקבל רק על a וכל השאר לבור דוחה).

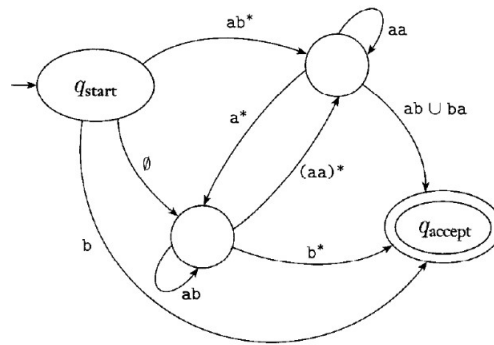
- אם $r = s \cup t$ אז קיימים A_s, A_t מה"א ומסגירות לאיחוד קיים אוטומט ל- $L(A_s) \cup L(A_t)$.

- אם $r = s \cdot t$ אז קיימים A_s, A_t מה"א ומסגירות לשרשר, קיים אוטומט ל- $L(A_s) \cdot L(A_t)$.

- אם $r = t^*$ אז מה"א יש A_t ששפתו שווה לשל t ולכן מסגירות לפעולת הכוכב, קיים אוטומט ל- $L(A_t)^*$.

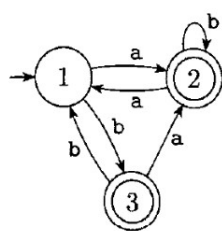
\Leftarrow : יהי A DFA ונוכיח שקיים לו ביטוי רגולרי r עם שפה שקולה. נוכיח בדוגמה של הרצת אלג' שמרדד DFA לביטוי רגולרי.

נניח שמותר לנו להשתמש ב-GNFA, שהוא NFA בעל קשתות עם ביטויים רגולריים במקום אותיות. בנוסף, נניח של- A (או ל-NFA המקביל לו) יש מצב התחלתי ומקבל יחיד (קל באמצעות צעדי אפסילון), וכן שהמצב ההתחלתי והמקבל זרים (גם קל עם צעדי אפסילון). ראו דוגמה ל-GNFA,

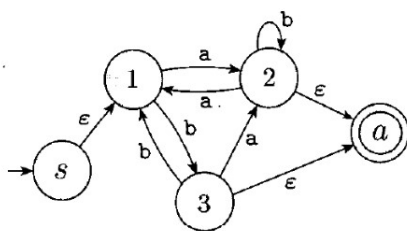


איור 22: GNFA לדוגמה, אפשר לעבור בין קשתות רק באמצעות מילה שעונה על הביטוי בקשת

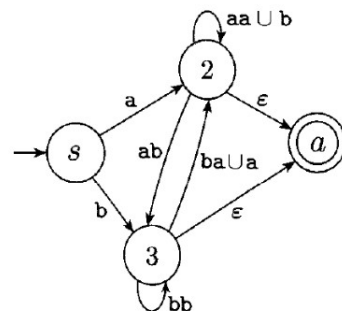
עתה נעקוב אחר הדוגמה שלקוחה מהתרגול כי אני לא מזוכיסט, ראו איור ואחריו הנחיה בנוגע למה אנחנו רואים.



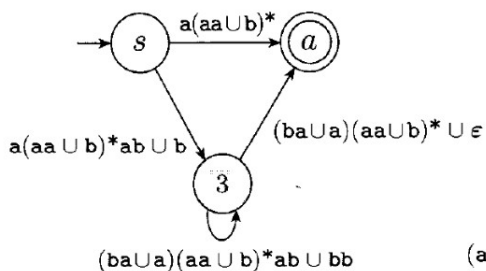
(a)



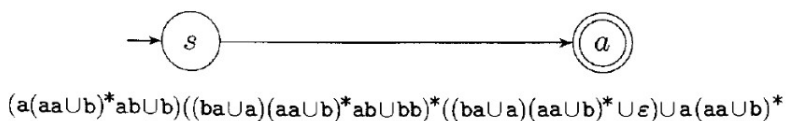
(b)



(c)



(d)



איור 23: GNFA לדוגמה, אפשר לעבור בין קשתות רק באמצעות מילה שעונה על הביטוי בקשת

במעבר הראשון אנחנו מוסיפים את המצב ההתחלתי והמקבל החדשים כדי לקיים את ההנחות שלנו.

במעברים הבאים אנחנו מוחקים מצבים (במקרה שלנו אחד כל פעם) ומחלצים מהם ביטויים רגולריים מתאימים עד שנישאר רק עם המצב ההתחלתי והמקבל החדשים. נציג נימוקים לכמה מהצמצומים האלה.

במעבר השני אנחנו מוחקים את 1:

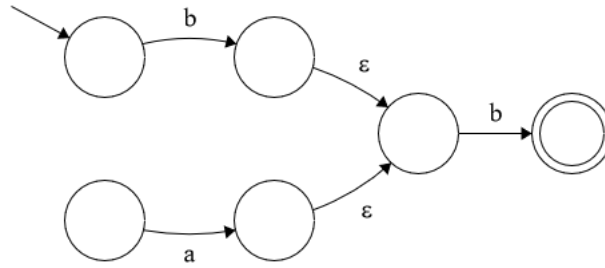
- ל-2 אפשר להגיע דרך 1 מ- s ולכן צמצמנו את צעד האפסילון;
- מ- s ל-3 צריך b ואז רצף כלשהו של bb , לכן יש לנו קשת b וחוג של 3 עם קשת bb ;
- כדי להגיע מ-3 ל-2 אפשר או ללכת ישר באמצעות a , או לעבור דרך 1 באמצעות b ואז a , כלומר $ba \cup a$;
- בנוסף, אפשר להגיע ל-2 מ-2 באמצעות סיבוב דרך 3 ו-1 ולכן יש לו חוג סביב עצמו עם ערך $aa \cup b$ מנימוק דומה לנ"ל.

במעבר השני אנחנו מוחקים את 2:

- מ- s ל- a אפשר להגיע או דרך 2 באמצעות a ואיזושהי כמות של סיבובים סביב 2 באמצעות $aa \cup b$.
- מ-3 ל- a אפשר להגיע עם מספר כלשהו של bb וזהו, או דרך 2 עם $ba \cup a$ ואז כמה סיבובי $aa \cup b$, או ישר עם אפסילון.

הרידור האחרון לא מורכב מדי, הוא די ישיר מבחינת האיחודים כי אין יותר מדי אפשרויות, רק לכתוב את זה זה נורא.

דוגמה $b \cdot (a \cup b)$, נוכל להרכיב אוטומט ל- $a, b, a \cup b$ ואז $b \cdot (a \cup b)$, זהו כל אחת מהבניות באיור השלם (שימו לב שב- $a \cup b$ שני השניים משמאל היו מקבלים אבל זה הוסר לטובת המצב הסופי).



איור 24 : NFA לביטוי הרגולרי הנ"ל

דוגמה תהי $L = \{1^{n^2} : n \in \mathbb{N}\}$. נראה ש- L לא רגולרית. נניח בשלילה ש- L רגולרית, לכן קיים קבוע ניפוח p כך שלכל מילה $w \in L$ עם $|w| > p$ ניתן לכתוב $w = xyz$ כך ש- $|xy| \leq p, |y| > 0$ ו- $xy^i z \in L$ לכל $i \geq 0$. נביט במילה $w = 1^{p^2} \in L$ אז $|w| > p$. נכתוב $w = xyz$ כאשר $x = 1^j, y = 1^k, z = 1^l$ ונבחר אותם כך ש- $k > 0$ ו- $k + j \leq p$. ננסה לנפח ב- $i = 2$. נשים לב כי

$$p^2 \stackrel{k \geq 0}{<} p^2 + k \stackrel{k+j \leq p}{\leq} p^2 + p < p^2 + 2p + 1 = (p+1)^2$$

כלומר $p^2 < |xy^2z| < (p+1)^2$ ולכן $xy^2z \notin L$ אינו ריבוע שלם ולכן הניפוח ב- $i = 2$ אינו ב- L , בסתירה לכך של- L רגולרית.

דוגמה $L = \{w \in \{0,1\}^* : \#_0 w = \#_1 w\}$. נראה כי L לא רגולרית. בהינתן p קבוע ניפוח, נבחר $w = 0^p 1^p$ ונכתוב $w = xyz$ כאשר $|xy| \leq p$ ו- $|y| > 0$ ולכן $x = 0^j, y = 0^k, z = 0^l 1^p$ כאשר $k > 0, j + k < p$. עבור $i = 2$, נקבל את הניפוח

$$xy^2z = 0^{j+2k} 0^l 1^p$$

וברור שיש יותר אפסים מאחדים ולכן הניפוח לא בשפה סתירה.

שבוע IV | משפט מייהיל-נרוד

הרצאה

חלק א' של ההרצאה

הגדרה $\forall L \subseteq \Sigma^*$, נגדיר יחס $\sim_L \subseteq \Sigma^* \times \Sigma^*$ כך שלכל $x, y \in \Sigma^*$, מתקיים $x \sim_L y$ אם ורק אם $x \cdot z \in L \iff y \cdot z \in L, \forall z \in \Sigma^*$.

הערה מילולית, $x \sim_L y$ אם לא משנה איזו מילה נדביק לסוף של שתיהן, הן או שתיהן יהיו בשפה או שתיהן לא.

דוגמה $L = (0+1)^* 0 (0+1)^*$. במקרה כזה $0 \sim_L 1$ כי $0 \cdot z \in L \iff 1 \cdot z \in L$ אבל $10 \notin L$ אבל $00 \in L$.

$\epsilon \sim_L 11 \sim_L 1$ כי $\forall z \in \Sigma^*, 11 \cdot z \in L \iff 1 \cdot z \in L$ אם ורק אם $1 \cdot z \in L$ (מילה היא בשפה אם האות הלפני אחרונה היא 0).

$10 \sim_L 01$ כי ϵ זנב מפריד (המילים עצמן מופרדות כבר).

טענה לכל שפה L , \sim_L היא יחס שקילות.

הוכחה: רפלקסיביות: $\forall x, x \sim_L x$.

סימטרי: $\forall x_1, x_2 \in \Sigma^*$, אם $x_1 \sim_L x_2$ אז $x_2 \sim_L x_1$ כי התנאי עצמו סימטרי.

טרנזיטיביות: $\forall x_1, x_2, x_3 \in \Sigma^*$ אם $x_1 \sim_L x_2$ וגם $x_2 \sim_L x_3$ מתקיים $x_1 \sim_L x_3$ כי אם בשלילה $x_1 \not\sim_L x_3$ קיים $z \in \Sigma^*$ כך ש- $x_3 \cdot z \notin L \iff x_1 \cdot z \in L$ (בה"כ על המספור), אבל

$$x_3 \cdot z \notin L \iff x_1 \cdot z \in L \iff x_2 \cdot z \in L \iff x_3 \cdot z \in L$$

■

סתירה.

הערה נסמן $[w]$ מחלקת השקילות של המילה w .

דוגמה עבור L הנ"ל, נמצא את מחלקות השקילות של היחס \sim_L .

ϵ ו-0 לא מקיימים את היחס, והמילה 1 מפרידה ביניהם. מתקיים $1 \in [\epsilon]$. 00 מחלקה חדשה, ומפורדת מהשתיים הראשונות ע"י ϵ .

01 גם מחלקה חדשה, וסה"כ המחלקות הן

$$[0] = 0, \Sigma^*10 \quad [\epsilon] = \epsilon, 1, \Sigma^*11 \quad [00] = \Sigma^*00 \quad [01] = \Sigma^*01$$

הערה נשים לב כי אם $x_1 \sim_L x_2$ וגם $x_3 \sim_L x_4$ אז $x_1 \sim_L x_3$ ו- $x_2 \sim_L x_4$ מפריד בין x_1 ו- x_3 אז x_2 ו- x_4 .

ניתן לראות זאת בדוגמה הנ"ל עבור $10 \sim_L 0$ ו- $101 \sim_L 01$ ו- ϵ , אך מתקיים ש- $10 \not\sim_L 01$ בין היתר בזכות ϵ .

משפט (מייהל-נרוד) $\forall L \subseteq \Sigma^*$, אזי $L \in \text{REG}$ אם ורק אם יש ל- \sim_L מספר סופי של מחלקות שקילות.

הוכחה: \Rightarrow נניח של- \sim_L יש מספר סופי של חלקות שקילות. נגדיר DFA $A = \langle \Sigma, Q, q_0, \delta, F \rangle$ שעבורו $L(A) = L$ נבחר

• Q מחלקות השקילות של \sim_L .

• $q_0 = [\epsilon]$.

• $\delta([w], \sigma) = [w \cdot \sigma]$.

• $F = \{[w] : w \in L\}$.

נשים לב שהגדרה של δ, F לא תלויה בבחירת הנציג (w) כי הרבה מאוד מצבים הם בעלי אותו נציג (אם $y \sim_L w$ אז $[y\sigma] = [w\sigma]$ $\forall z$ כי אחרת σz מפריד של (y, w)).

נוכיח שלכל $w \in \Sigma^*$, $\delta^*(q_0, w) = [w]$ ולכן מהגדרת F, L אם $w \in L$ אם $\delta^*(q_0, w) = F$ ונסיים. באינדוקציה על $|w|$.

בסיס $\delta^*(q_0, \epsilon) = q_0 = [\epsilon]$ ואכן $w = \epsilon : (w = \epsilon)$.

צעד $(|w| \rightarrow |w| + 1)$:

$$\delta^*(q_0, u \cdot \sigma) = \delta(\delta^*(q_0, u), \sigma)$$

$$\stackrel{\text{ה"נ}}{=} \delta([u], \sigma)$$

$$\stackrel{\text{הגדרה}}{=} [u\sigma]$$

\Leftarrow נניח ש- $A = \langle \Sigma, Q, q_0, \delta, F \rangle$ הוא DFA שמזוז את L ונראה של- \sim_L מספר סופי של מחלקות שקילות. נחסום את המספר הזה עם מספר המצבים ונסיים.

נגדיר יחס $\sim_A \subseteq \Sigma^* \times \Sigma^*$ ונאמר כי $x, y \in \Sigma^*$ מקיימות $x \sim_A y$ אם $\delta^*(q_0, x) = \delta^*(q_0, y)$.

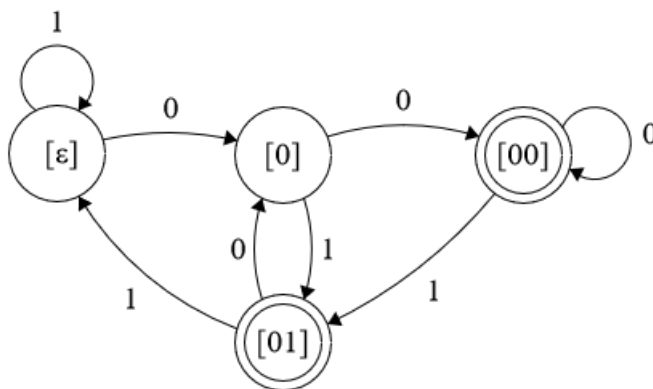
אם $x \sim_A y$ אז אין להן זנב מפריד כי xz, yz נפגשות אחרי x, y ומשם ממשיכות יחד בריצה על z ולכן תמיד יגיעו לאותו המקום, ולכן $x \sim_L y$ ופורמלית, אם $x \sim_A y$ אז $\forall z \in \Sigma^*$,

$$\delta^*(q_0, xz) = \delta^*(\delta^*(q_0, x), z) = \delta^*(\delta^*(q_0, y), z) = \delta^*(q_0, yz)$$

ולכן $xz \in L \iff yz \in L$ ולכן $x \sim_L y$.

מכאן שמספר מחלקות השקילות של \sim_A חוסם את מספר מחלקות השקילות של \sim_L , והראשון חסום ע"י $|Q|$ ולכן גם האחרון ולכן הוא סופי. ■

דוגמה נפעיל את המשפט על הדוגמה הנ"ל $L = (0 + 1)^* 0 (0 + 1)$,



איור 25: אוטומט שמתאים לשפה L

כאשר בנינו כל קשת ע"י בדיקה של היכן נמצא הנציג יחד עם האות על הקשת, לדוגמה [01] עם 0 הולך ל-0 כי 010 הוא במחלקת השקילות של 0, ושאר הקשתות בהתאם.

שימושים של משפט MN

1. סיווג ל-REG או לא REG.

דוגמה $L = \{0^n 1^n : n \geq 0\}$ אינה רגולרית כי $0^i \approx 0^j, \forall i \neq j \geq 0$ כי 1^i זנב מפריד $0^i 1^i \in L$ אבל $0^j 1^j$ לא) ולכן יש אינסוף מחלקות שקילות ל- \sim_L וסיימנו.

דוגמה $L = \{0^i 1^j : \gcd(i, j) \neq 1\}$. נראה שעבור שני ראשוניים, $p_1 \neq p_2$, $0^{p_1} \approx_L 0^{p_2}$ (כאן $j = 0$). נשים לב כי 1^{p_1} הוא זנב מפריד (כי $0^{p_1} 1^{p_1} \in L$ אבל $0^{p_2} 1^{p_1}$ לא). לכן ל- \sim_L אין סוף.

2. צמצום/מזעור DFA-ים.

הרעיון הוא שאם לאוטומט יש יותר מצבים ממחלקות שקילות ל- \sim_L , אפשר לצמצם את ה-DFA עוד. נראה אלג' שבהנתן DFA $A = \langle \Sigma, Q, q_0, \delta, F \rangle$, נחזיר DFA A' שקול ל- A כך שלכל DFA A'' , אם $L(A'') = L(A)$, אז $|A'| \leq |A''|$ (DFA A' מינימלי עבור $L(A)$). מעבר לכך, נראה שאוטומט זה הוא יחיד עד כדי שמות.

מזעור אוטומטים

נגדיר סדרה של יחסים \sim_i על $Q \times Q$, $\forall i \geq 0$. הרעיון הוא ש- $s_1 \sim_i s_2$ אם $\forall z \in \Sigma^*, |z| \leq i$, $\delta^*(s_1, z) \in F \iff \delta^*(s_2, z) \in F$. כלומר אינטואיטיבית $s_1 \sim_i s_2$ אם s_1, s_2 מסכימות על אילו מילים עד אורך i מתקבלות (כשהריצה מתחילה מהן).

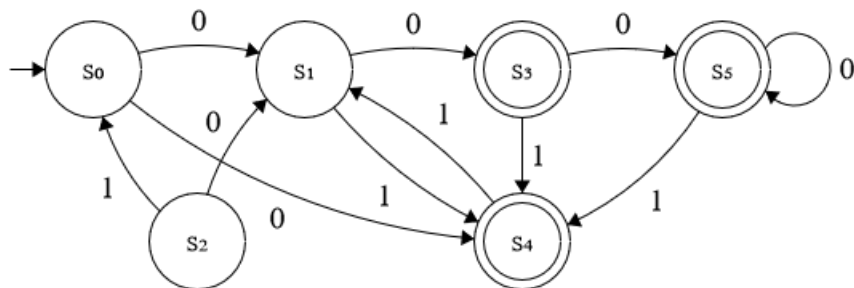
ככל ש- i יותר גדול, כך יש יותר מילים שצריך שהתנאי הזה יתקיים עליהן ולכן מחלקות השקילות שלו יגדלו (ומספרן יגדל). מתישהו נפסיק לעדן את מחלקות השקילות ומחלקות השקילות שנקבל יספקו לנו את המצבים ל-DFA המינימלי.

הגדרה נגדיר את הסדרה \sim_i באופן אינדוקטיבי.

בסיס ($i = 0$): $s_1 \sim_0 s_2$ אם $s_1 \in F \iff s_2 \in F$ (ויש לו שתי מחלקות שקילות, כל המקבלים וכל הלא מקבלים).

צעד ($i \rightarrow i + 1$): נגדיר $s_1 \sim_{i+1} s_2$ אם $s_1 \sim_i s_2$ וגם $\forall \sigma \in \Sigma^*, \delta(s_1, \sigma) \sim_i \delta(s_2, \sigma)$ (כלומר אם s_1, s_2 מסכימים על מילים באורך i וגם על כל הארכה באורך 1).

דוגמה נביט באוטומט הבא שמזהה את השפה $L = (0 + 1)^* 0 (0 + 1)$



איור 26: אוטומט שמתאים לשפה L

עבור \sim_0 , מחלקות השקילות שלנו הן

$$\{\{s_0, s_1, s_2\}, \{s_3, s_4, s_5\}\}$$

עבור מילים באורך 1, נעדן את מחלקת השקילות. האם $s_0 \sim_1 s_1$? מתקיים $s_0 \sim_0 s_1$ אבל $s_0 \sim_1 s_1$ ולכן התשובה היא לא. עם זאת $s_0 \sim_1 s_2$ כן מתקיים כי הפעלה של 0 ו-1 מובילות אותנו למצבים שהם באותה מחלקת שקילות בהתאמה. אחרי חישוב מקבלים שמחלקות השקילות ל- \sim_1 הן

$$\{\{s_0, s_1\}, \{s_2\}, \{s_3, s_5\}, \{s_4\}\}$$

וזהו עבור \sim_2 מקבלים את אותה מחלקת שקילות ושם נעצור (הגענו לנקודת שבת) ואכן ארבעת המחלקות הללו נותנות לנו אוטומט מזערי.

חלק ב' של ההרצאה

נביט בסדרת היחסים שהגדרנו $\{\sim_i\}$ (שכל אחד מהם אוסף זוגות של מצבים). בהכרח שעבור i גדול מספיר, נקבל $\sim_i = \sim_{i+1}$ (שוויון בין קבוצות המוכלות ב- $Q \times Q$) כי אם $s_1 \sim_i s_2$ אז $s_1 \sim_{i+1} s_2$ ולכן $\sim_i \subseteq \sim_{i+1}$.

מכאן שאו שהגענו לנקודת שבת ונעצור, או שהורדנו לפחות זוג אחד מ- i , ולכן תוך לכל היותר $|Q|^2$ איטרציות נעצור.

בנוסף, המעבר מ- \sim_i ל- \sim_{i+1} מתבצע בזמן פולינומיאלי, שכן יש מספר פולינומיאלי של זוגות (לכל היותר $|Q|^2$) וחישוב האם זוג עובר ליחס הבא או לא דורש זמן קבוע.

טענה לכל $i \geq 0$, $s_1, s_2 \in Q$, מתקיים $s_1 \sim_i s_2$ אם ורק אם לכל מילה w באורך i , $\delta^*(s_1, w) \in F \iff \delta^*(s_2, w) \in F$.

הערה בתרגיל נוכיח שזה מספיק כדי להראות שמחלקות השקילות מהוות מצבים לאוטומט המזערי.

הוכחה: נראה באינדוקציה על i :

בסיס $w = \epsilon : (i = 0)$ מההגדרה $s_1 \sim_0 s_2 \iff (\delta^*(s_1, \epsilon) = s_1 \in F \iff \delta^*(s_2, \epsilon) = s_2 \in F)$

$$: (i \rightarrow i + 1) \text{ טעם}$$

⇐ נניח ש- $s_1 \sim_{i+1} s_2$ נוכיח שלכל מילה w , אם $|w| \geq i+1$ אז $\delta^*(s_1, w) = \delta^*(s_2, w) \in F$ תהי w כנ"ל.

• אם $|w| \geq i$ ולכן $s_1 \sim_{i+1} s_2$ ולכן מה"א הטענה מתקיימת עבור w .

- אם $i = |w| + 1$ אז $w = \sigma y$ עבור $\sigma \in \Sigma, y \in \Sigma^*$ ומהגדרת \sim_{i+1} ,

$$s'_1 = \delta(s_1, \sigma) \sim_i \delta(s_2, \sigma) = s'_2$$

ולכן מה"א (עבור y שהיא באורך i)

$$\delta^*(s'_1, y) \in F \iff \delta^*(s'_2, y) \in F$$

ולכן

$$\delta^*(s_1, \sigma y) = \delta^*(\delta(s_1, \sigma), y) \in F \stackrel{\text{הביטוי הנ"ל}}{\iff} \delta^*(\delta(s_2, \sigma), y) \in F = \delta^*(s_2, \sigma y)$$

כלומר w מקיימת את התנאי.

\Rightarrow : נניח ש- s_1, s_2 מסכימים מילים עד לאורך $i + 1$ ונוכיח ש- $s_1 \sim_{i+1} s_2$.

נניח בשלילה ש- $s_1 \not\sim_{i+1} s_2$. לכן או ש- $s_1 \not\sim_i s_2$ או שקיימת $\sigma \in \Sigma$ כך ש- $\delta(s_1, \sigma) \not\sim_i \delta(s_2, \sigma)$ (מההגדרה).

אם $s_1 \not\sim_i s_2$, קיימת מילה y באורך $i \geq$ כך ש- $\delta(s_1, y) \not\sim_i \delta(s_2, y)$ כלומר s_1, s_2 לא מסכימים על מילה באורך i סתירה.

אם קיימת σ כך ש- $\delta(s_1, \sigma) \not\sim_i \delta(s_2, \sigma)$, אז $\delta(s_1, \sigma), \delta(s_2, \sigma)$ הם מצבים לא ביחס \sim_i ולכן מה"א הם לא מסכימים על השפה עד אורך $i \geq$.

כלומר, קיימת y עם $|y| \geq i$ כך ש- $\delta^*(\delta(s_1, \sigma), y) \in F$ אבל $\delta^*(\delta(s_2, \sigma), y) \notin F$ בסתירה לכך ש- s_1, s_2 מסכימים על מילים באורך $i + 1$.

■

תרגול

טענה תהי $f : \mathbb{N} \rightarrow \mathbb{N}$ מונוטונית עולה ממש כך ש- $f(n)$ היא $\omega(n)$, אזי השפה $L_f = \{a^{f(n)} : n \in \mathbb{N}\}$ לא רגולרית.

הוכחה: נשתמש בלמת הניפוח ע"י מציאה לכל n , מילה באורך בין $f(n)$ ל- $f(n+1)$, ונסיים.

טענת עזר תהי f כנ"ל. אזי $\forall K, N \in \mathbb{N}$, קיים $n > N$ כך ש- $f(n+1) - f(n) > K$ (כלומר נצליח לחסום ממתחת את ההפרשים בין האיברים, עבור מספרים מספיק גדולים).

הוכחה: (של טענת העזר) נניח בשלילה שלא קיים, לכן קיימים $K, N \in \mathbb{N}$ שעבורם $f(n+1) - f(n) \leq K, \forall n > N$. בפרט, קיים $M \in \mathbb{N}$ כך ש- $M - f(n) \leq f(n+1) - f(n)$ (מקסימום ההפרשים עד N).

לכן מתקיים $f(2) - f(1) \leq M$, ואז $f(3) \leq f(2) + M \leq f(1) + 2M$ ובצעד ה- n , $f(n) \leq (n-1)M + f(1)$ ולכן

$$\frac{f(n)}{n} \leq \frac{n-1}{n}M + \frac{f(1)}{n} \xrightarrow{n \rightarrow \infty} M + 0$$

ולכן ממונוטוניות הגבול, $\lim_{n \rightarrow \infty} \frac{f(n)}{n} \leq M$ בסתירה לכך שהגבול הזה הוא ∞ מהגדרת ω .

■

נחזור לטענה. נניח בשלילה שלמת הניפוח מתקיימת ויהי $p > 0$ קבוע הניפוח. נבחר $N = K = p$ ונביט במילה $a^{f(n)}$ עבור $n > p$ שמטענת העזר, מקיים $f(n+1) - f(n) > k$. נבחר $w = xyz = a^l a^m a^s$ כאשר $l + m + s = f(n)$ ו- $0 < m$. נביט ב- xy^2z . $|xy^2z| = f(n) + m$ ומתקיים

$$f(n) < f(n) + m \leq f(n) + p < f(n+1)$$

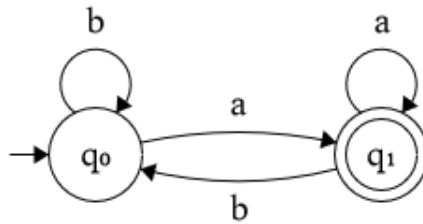
כאשר המעבר הראשון והשני נובעים מהתנאים של למת הניפוח על l, m והמעבר השלישי נובע מתטענת העזר ($p = K$). לכן $xy^2z \notin L$. בסתירה ללמת הניפוח.

למעשה המעבר המהותי הוא שניפוחו ב- m את המילה, אבל m קטן מאשר החסם התחתון שמצאנו להפרש ולכן הוא לא במילה. ■

הערה בכתובה מתמטית נטו, יחס השקילות שמוגדר במייהיל-נרוד מוגדר באופן הבא,

$$\forall x, y \in \Sigma^* (x \sim_L y \iff (\forall z \in \Sigma^*, xz \in L \iff yz \in L))$$

דוגמה נביט בשפה $\{w \in \Sigma \{a, b\}^* : w \text{ מסתיימת ב-} a\}$. L היא רגולרית כי האוטומט הבא מזהה אותה.



איור 27: אוטומט שמתאים לשפה L

נוכיח זאת עם MN. נסתכל על המילים בשפה באופן שיטתי.

- אם $x, y \in \Sigma^*$ מסתיימות ב- a : $xz \in L$ אם $z = \epsilon$ או ש- z עצמה מסתיימת ב- a (וזה אותו התנאי על y) ולכן $x \sim_L y$ וזו מחלקת שקילות אחת.
- אם $x, y \in \Sigma^*$ לא מסתיימות ב- a : $xz \in L$ אם z מסתיימת ב- a (באותו האופן על y) ולכן $x \sim_L y$ וזו עוד קבוצה במחלקת שקילות, עדיין לא ידוע אם שונה מהקודמת.
- אם x לא מסתיימת ב- a ו- y מסתיימת ב- a : $x \sim_L y$ כי $z = \epsilon$ זנב מפריד, ולכן שתי מחלקות השקילות הנ"ל שונות ומיפינו את כל המרחב לשתי מחלקות שקילות.

דוגמה נביט בשפה $\{w \text{ מאורך שאינו חזקה של } 2 : w\}$. $L = \{a\}$. מעל $\Sigma = \{a\}$. ראינו שהמשלימה של השפה הזו היא לא רגולרית ולכן נצפה שגם זו תהיה (אחרת נוכל לבנות אוטומט עם מצבים מקבלים הפוכים).

נראה ש- L לא רגולרית ע"י מציאת אינסוף מח"ש ל- \sim_L . יהיו $m > n \in \mathbb{N}$. נביט במילים $x = a^{2^n}, y = a^{2^m}$ שתיהן ב- Σ^* (ולא ב- L אבל זה לא מעניין). נשים לב שעבור $z = a^{2^n}$ נקבל ש- $xz = a^{2^{m+1}} \in L$ אבל $yz = a^{2^n(2^{m-n}+1)} \notin L$ ולכן $x \not\sim_L y$. כלומר, לכל $x, y, m > n$ כנ"ל מצאות במח"ש שונות ויש אינסוף זוגות מספרים כאלה ולכן יש ∞ מח"ש ל- \sim_L .

תרגיל יהי DFA $A = \langle Q, \{0, 1\}, q_0, \delta, F \rangle$ עם $|Q| = r$. נתון ש- $0^r 1^r \in L(A)$. מה מהבאים נכון בהכרח?

1. $0^* 1^* \subseteq L(A)$

2. $L(A) \subseteq 0^* 1^*$

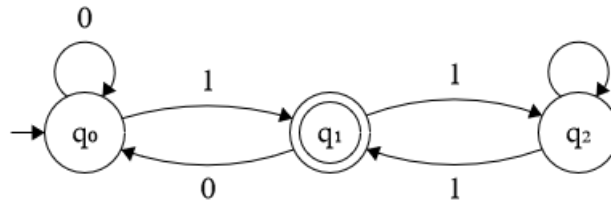
3. (1) לא בהכרח נכון אבל לכל $i \geq 1$ מתקיים $0^{ir} 1^{ir} \in L(A)$

4. (1) לא בהכרח נכון אבל קיים $k \geq 1$ שעבורו לכל $i \geq 1$, $0^{r+ik} 1^{r+k} \in L(A)$

פתרון (1) לא נכון כי אוטומט עם שני מצבים לשפה שמכילה את כל המילים עם מספר זוגי של אפסים. אוטומט כזה יקבל את 0011 אבל לא את 00111.

(2) לא נכון כי עבור האוטומט הנ"ל, $010 \in L(A)$ (מספר זוגי של אפסים) אבל $010 \notin 0^* 1^*$.

(3) לא נכון כי עבור האוטומט באיור, שעבורו $r = 3$, מתקיים $0^3 1^3 \in L(A)$ אבל $0^6 1^6 \notin L(A)$. זה משום $0^3 1^3$ יגיע עד ל- q_2 ויחזור ל- q_1 , ואילו $0^6 1^6$ יגיע עד ל- q_2 וילך הלך ושוב ויסיים ב- q_2 ולא יקבל.

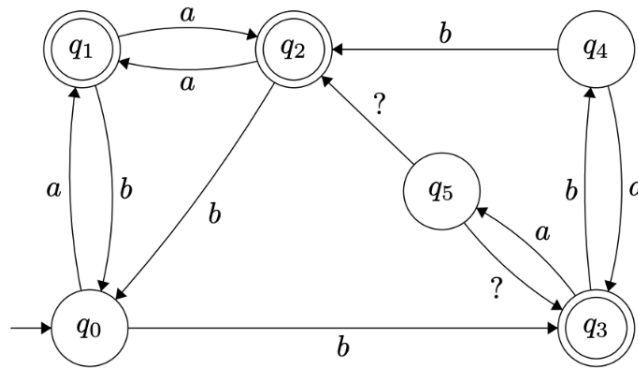


איור 28: אוטומט שמפריך את (3)

(4) כן נכון, נוכיח עם למת הניפוח. בריצות על המילים 0^r ו- 1^r , יש מצב שחוזר על עצמו (לפחות אחד) ולכן בריצה על 0^r יש מעגל באורך k_1 ובריצה על 1^r יש מעגל באורך k_2 .

את המעגלים האלה נוכל לשכפל עוד ועוד ולהגיע לאותו המצב. נביט בריצה של A הדומה לזו על $0^r 1^r$ אבל רצה על המעגל של 0^{k_2} פעמים ועל המעגל של 1^{k_1} פעמים. נסמן $k = k_1 k_2$ ואז הטענה מתקבלת (הוספנו ל- 0^r אפסים וגם ל- 1^{k_1} אחדות ונשארו בשפה).

תרגיל נתון ה-DFA כבאיור, A ,



איור 29: האוטומט A לתרגיל

נתון כי ל- $L(A)$ יש 4 מח"ש MN. מה אמור להיות במקום סימני השאלה באיור?

$$1. \delta(q_5, a) = q_2, \delta(q_5, b) = q_3$$

$$2. \delta(q_5, a) = q_5, \delta(q_5, b) = q_2$$

3. (1) ו-(2) נכונות.

4. כל התשובות לא נכונות.

פתרון נשים לב שמכך של- A 4 מח"ש MN, האלג' המצמצם אמור לאחד מצבים עד שנגיע ל-4.

• הצעד הראשון באלג' יגיע למח"ש (לפי F) $\{q_1, q_2, q_3\}, \{q_0, q_4, q_5\}$.

• השלב השני יפריד עם זנב a לפחות ל- $\{q_3\}, \{q_1, q_2\}$ ו- $\{q_0, q_4, q_5\}$ עבור תשובות (1), (2).

• בשלב השלישי, ab לא יפריד לנו שום דבר. עבור זנב aa , נקבל $\{q_1, q_2\}, \{q_3\}, \{q_0, q_5\}, \{q_4\}$ כי q_4 מגיע למצב לא מקבל ו- q_0, q_5 למקבל (בשתי התשובות).

כאן נעצור כי הגענו ל-4 מחלקות שקילות ובגלל שתשובות (1), (2) מקיימות את המח"ש האלה, אלה התשובות הנכונות ולכן (3) היא התשובה הנכונה.

הערה נשים לב שאם היינו בוחרים זנבות אחרים, היינו יכולים להגיע למחלקות שקילות אחרות אבל עדיין מספר מחלקות שקילות שווה.

שבוע V | שפות חסרות הקשר

הרצאה

חלק א' של ההרצאה

הערה בהרצאה הקודמת ראינו איך למזער אוטומט דטרמיניסטי. כיצד נוכל למזער NFA? אם אנחנו יודעים להכריע האם קיים NFA שקול עם k מצבים, נוכל לעבור על כל $k \in \mathbb{N}$ עד שנענה כן וזה יהיה NFA מינימלי.

בעיית הריקנות

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

אפשר להכריע ע"י חיפוש בגרף (DFS/BFS) החל מ- Q_0 , ואם מגיעים לקודקוד כלשהו והוא מקבל נחזיר "שקר" ואחרת אם כל הקודקודים הישיגים לא מקבלים, נחזיר "אמת".

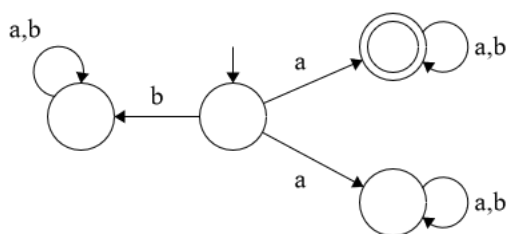
• הבעיה הדואלית לבעיית הריקנות, בעיית האוניברסליות, שואלת, בהינתן אוטומט A , האם $L(A) = \Sigma^*$?

מתקיים $L(A) = \Sigma^*$ אם ורק אם $\overline{L(A)} = \emptyset$ (כאשר \overline{A} הוא המשלים של A , נגדיר אותו עוד רגע). לכן מספיק שנייצר את \overline{A} ונבדוק האם $L(\overline{A}) = \emptyset$.

בניית המשלים של A

• עבור A DFA: \overline{A} הוא אותו אוטומט עם מצבים מקבלים הכל חוץ מהמצבים המקבלים של A ($F' = Q \setminus F$).

• עבור A NFA: החלפת המצבים המקבלים לא מספיקה, לדוגמה באיור הבא, נקבל גם במקורי וגם בבנייה החדשה שמילים שמתחילות ב- a מתקבלות. הבעיה היא שכאן הבנייה מקבלת את כל המילים שקיימת להן ריצה לא מקבלת, ולא כזו שכל ריצה שלהן היא לא מקבלת.



איור 30: אוטומט שסותר את הבניה המקורית למשלים

מה שכן יעבוד, הוא לעשות דטרמיניזציה ל-DFA שקול A' , דואליזציה למשלים $\overline{A'}$ ובדיקת ריקנות ל- $\overline{A'}$.

הסיבוכיות של אלג' זה היא אקספוננציאלית כי A' במקרה הגרוע הוא בעל מספר מצבים אקספוננציאלי ב- n גודל ה-NFA.

משפט אין פולינום p כך שבהינתן (כל) NFA A , ניתן לייצר \overline{A} עם $|\overline{A}| \leq p(|A|)$.

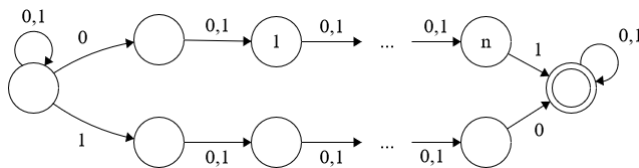
מסקנה האלג' שהראנו למעלה הוא הכי טוב שאפשר ואין אחד עם סיבוכיות קטנה יותר, כי זה בכל מקרה אקספוננציאלי.

הוכחה: נראה משפחה של שפות $\{L_n\}_{n=1}^{\infty}$ כך שלכל n קיים NFA A_n עבור L_n עם $\mathcal{O}(n^2)$ מצבים ולכל NFA $\overline{A_n}$ עבור $\overline{L_n}$ דורש לפחות 2^n מצבים.

נבחר $\overline{L_n} = \{ww : w \in \{0+1\}^n\}$ כך שלדוגמה $\overline{L_2} = \{0000, 0101, 1010, 1111\}$ (כל השאר המילים).

נראה שקיים NFA עם $\mathcal{O}(n^2)$ מצבים ל- $\overline{L_n}$.

נשים לב כי $w \in L_n$ אם $|w| \neq 2n$ או שקיים אינדקס $i \in [n]$ כך ש- $w_i \neq w_{n+i}$. לכן האוטומט הבא יזהה נכון את $\overline{L_n}$, כי הוא יכול לנחש כל אינדקס לא נכון (נניח שניחש שזה 0 בהתחלה ו-1 בסוף אז אחרי n צעדים במסלול העליון הוא יקבל).



איור 31: אוטומט שמזהה את L_n

זהו אוטומט עם $\mathcal{O}(n)$ מצבים.

נראה שכל NFA ל- $\overline{L_n}$ צריך לפחות 2^n מצבים. נניח בשלילה שקיים NFA המזהה את $\overline{L_n}$ עם פחות מ- 2^n מצבים.

לכל מילה $u \in (0+1)^n$, נתבונן בקבוצת המצבים

$$good(u) = \{s : u \text{ קריאת } s\text{-ב-} s \text{ שמבקרת על } \overline{A_n} \text{ של } uu\} \subseteq Q$$

כלומר אוסף המצבים שבהם בדיוק באמצע ריצה מקבלת על uu הגענו אליהם.

מהיות מספר המצבים של $\overline{A_n}$ פחות מ- 2^n מצבים, ולכן מעקרון שובך היונים קיימים $u_1 \neq u_2 \in (0+1)^n$ כך ש- $good(u_1) \cap good(u_2) \neq \emptyset$

ולכן אם s בחיתוך הזה, מתקיים

$$s \in \delta^*(Q_0, u_1), \quad \delta^*(s, u_2) \cap F \neq \emptyset \quad \Rightarrow \quad \delta^*(Q_0, u_1 u_2) \cap F \neq \emptyset$$

כלומר $\overline{A_n}$ קיבל את $u_1 u_2$ בסתירה לכך שברור ש- $u_1 u_2 \in \overline{L_n}$ (היא השפה עם כל המילים שאינן חזרה על מילה).

■

שפות חסר הקשר

שפות חסרות הקשר מוגדרות ע"י דקדוק חסר הקשר.

דוגמה נביט בדקדוק הבא,

$$A \rightarrow 0A1$$

$$A \rightarrow B$$

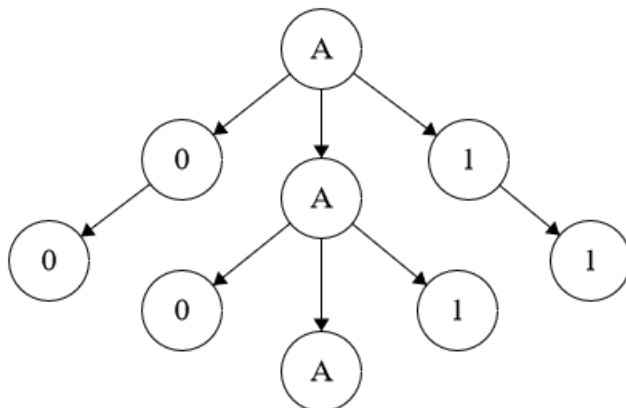
$$B \rightarrow \#$$

נשים לב שיש לנו משתנים A, B , טרמינלים (א"ב) $0, 1, \#$, חוקים (המשוואות) ומשתנה התחלתי.

במקרה כזה שרשרת פעולות הגזירה הבאה מייצרת לנו מילה, $A \rightarrow 0A1 \rightarrow 00A11 \rightarrow 00B11 \rightarrow 00\#11$, כאשר נשים לב שהשפה

$$L(G) = \{0^n \# 1^n\}$$

נוכל בנוסף לצייר את עץ הגזירה של הריצה הנ"ל, כאשר העלים של העץ משמאל לימין מייצרים לנו את המילה הסופית שקיבלנו בשרשרת הגזירה



איור 32 : עץ הגזירה של שרשרת הגזירות הנ"ל

הערה דקדוק חסר התחיל מעיבוד שפות טבעיות, שם נוכל לאפיין תארים ושמות עצם על ידי גזירה, לדוגמה $N \rightarrow AN$ מאפשר הוספת שם תואר לשם עצם באנגלית.

הגדרה דקדוק חסר הקשר הוא $G = \langle V, \Sigma, R, S \rangle$ כאשר :

- V קבוצה סופית של משתנים.
- Σ קבוצה סופית של אותיות.
- R קבוצה של חוקי גזירה מהצורה $V \rightarrow (V \cup \Sigma)^*$.
- $S \in V$ משתנה התחלתי.

הערה הדקדוק נקרא חסר הקשר כי יש בצד שמאל רק משתנה והוא יחיד.

הגדרה אם $A \rightarrow w$ ו- $w, u, v \in (V \cup \Sigma)^*$ אז יצירה/גזירה היא המעבר $vAu \Rightarrow vwu$.

אם $u, w \in (V \cup \Sigma)^*$ אז $u \Rightarrow^* w$ אם קיים $k \geq 1$ ו- u_1, \dots, u_k כך ש- $(V \cup \Sigma)^*$ $u = u_1 \Rightarrow \dots \Rightarrow u_k = w$ (ניתן לגזור מ- u ולהגיע ל- w).

הגדרה עבור G דקדוק ח"ה, נגדיר את השפה שלו להיות $L(G) = \{w : w \in \Sigma^* \wedge S \Rightarrow^* w\}$. שפה L היא שפה ח"ה (ונסמן $L \in CFL$) אם יש G CFG כך ש- $L = L(G)$.

דוגמאות

1. $G = \langle \{S, A\}, \{0, 1\}, R, S \rangle$ כאשר החוקים הם

$$S \rightarrow A1A$$

$$A \rightarrow \epsilon | 0A | 1A$$

כל תהליך גזירה יתחיל ב- $A1A$. $S \rightarrow$ לכן $\epsilon \notin L(G)$ אבל $1 \in L(G)$. קל לראות ש- $L(G) = (0+1)^* 1 (0+1)^*$.

2.

$$S \rightarrow 0S1|SS|\epsilon$$

מגדיר את שפה הסוגריים המקוננים חוקית כאשר 0 הוא (ו-1 הוא). נצטרך שבכל רישא לא יהיו יותר 1-ים מ-0-ים, ובסוף יהיו לנו מספר שווה של 0-ים ו-1-ים.

נשים לב כי $L(G) \cap \{0^*1^*\} = \{0^n1^n : n \geq 0\}$ ולכן $L(G)$ לא רגולרית.

משפט $\text{REG} \subseteq \text{CFL}$.

הוכחה: בהינתן $\text{DFA } A = \langle Q, \Sigma, q_0, \delta, F \rangle$, נבנה $G = \langle V, \Sigma, R, S \rangle$ כך ש- $L(G) = L(A)$. נבחר

$$V = \{V_q : q \in Q\} \bullet$$

$$S = V_{q_0} \bullet$$

• לכל מצב $q \in Q$ ו- $\sigma \in \Sigma$, נסמן $s = \delta(q, \sigma)$, נוסיף חוק $V_q \rightarrow \sigma V_s$. אם $q \in F$, נוסיף (מעבר לנ"ל), $V_q \rightarrow \epsilon$.

הרעיון כאן הוא שגזירה של מילה מ- V_q תסתיים בדיוק על כל המילים שמתקבלות מ- q .

נראה שלכל מצב $q \in Q$ ומילה w , $\delta^*(q, w) \in F \iff V_q \Rightarrow^* w$. נסמן $w = \sigma_1 \dots \sigma_k$ ולכן $\delta^*(q, w) \in F$ אם יש ריצה r_0, \dots, r_k של A על w כך ש- $r_0 = q$ ולכל $0 \leq i \leq k$, $r_{i+1} = \delta(r_i, \sigma_{i+1})$, אם "ם

$$V_{r_0} \Rightarrow \sigma_1 V_{r_1} \Rightarrow \dots \Rightarrow \sigma_1 \dots \sigma_k V_{r_k} \Rightarrow \sigma_1 \dots \sigma_k \epsilon = w$$

■

חלק ב' של ההרצאה

דוגמה הדקדוק $S \rightarrow aSa|bSb|\epsilon$ מייצר פולינדרומים באורך זוגי (באינדוקציה הוא מוסיף בכל צד אות). אם נרצה כל אורך, נוסיף $a|b$ לחוק היחיד שלנו.

משפט (למת הניפוח ל-CFL) תהי $L \in \text{CFL}$ אזי קיים $p \geq 0$ קבוע הניפוח כך שלכל מילה $w \in L$ עם $|w| \geq p$, ניתן לכתוב $w = uvxyz$ כאשר מתקיים

$$1. |vxy| \leq p$$

$$2. |vy| > 0$$

$$3. \text{לכל } i \in \mathbb{N}_0, uv^i xy^i z \in L$$

דוגמה שפת הפלינדרומים מקיימת את למת הניפוח. נבחר $p = 3$, ואז אם $w \in L$ ו- $w \geq 3$:

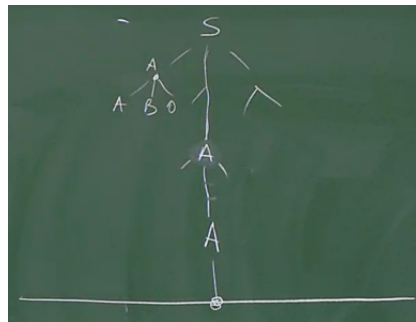
• אם $|w|$ אי זוגי, נבחר x להיות האות האמצעית, y, z שכנותיה ו- u הרישא והסיפא בהתאמה.

• אם $|w|$ זוגי נבחר x להיות ϵ וכל השאר כנ"ל (v, y שמאל לאמצע וימין לאמצע בהתאמה).

רק כך המילה המנופחת תהיה מספיק גדולה גם ב- $i = 0$, בחלוקות אחרות לא היו לנו מספיק אותיות בניפוח $i = 0$.

הוכחה: בלמת הניפוח ל-REG מצאנו מעגל במצבים (כאשר $|Q| > p$) וחזרנו עליו i פעמים. יהי $b \geq 2$ האורך של צד ימין ארוך ביותר בדקדוק של L (כלומר ב- $(2) \rightarrow (1)$ מדובר במספר המשתנים/טרמינלים ב- (2)).

עתה נבחר p שיבטיח שבעץ הגזירה של מילים באורך גדול מ- p , יש מסלול עם משתנה שחוזר לפחות פעמיים (כבאיור, A יכול לחזור אין ספור פעמים, לא תמיד מיד אחרי עצמו).



איור 33: עץ הגזירה עם חזרה של משתנה

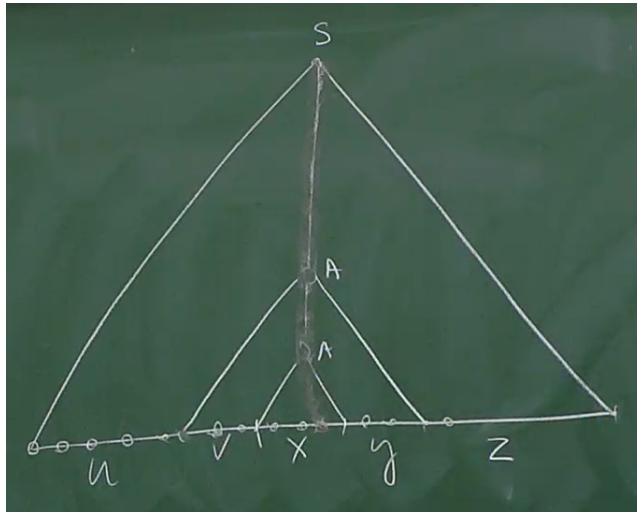
מתקיים שדרגת הפיצול של העץ (מספר הבנים של קודקוד כלשהו) $b \geq 2$ כי כל פיצול מתאים לחוק. נטען שהמילה מספיק ארוכה כדי שמשתנה יחזור פעמיים ועליו נוכל לחזור שוב ושוב.

נזכור כי אם מספר העלים $b^{|V|+1} \leq |V| + 1$ אז הגובה (ממבני נתונים).

נבחר $p = b^{|V|+1}$ ותהי $w \in L$ עם $|w| \geq p$. נתבונן בעץ הגזירה הקטן ביותר של w (יש כמה דרכים אולי לגזור ולהגיע ל- w). בהכרח מתקיים שגובה העץ $|V| + 1 \leq$ מההבחנה הנ"ל.

יש $|V|$ משתנים ויש עלה עם עומק $|V| + 1$ (מהגדרת הגובה) שמכיל רק משתנים (כי אם היה טרמינל היינו עוצרים ולא ממשיכים הלאה).

לכן יש משתנה (שמופיע כ- A באיור הבא) שחוזר על עצמו באחת מ- $|V| + 1$ הרמות הקרובות לעלים מעקרון שובך היונים.



איור 34 : חלוקה של המילה על פני עץ הגזירה, משולש מגדיר תת-עץ גזירה

נבחר חלוקה כבאיור הנ"ל (מספיק פורמלי). נראה שמתקיימים התנאים.

1. $|vxy| \leq b^{|V|+1}$ כי בחרנו את החזרה הכי נמוכה של A בעץ, שהיא בגובה (ביחס לעלים) לכל היותר $|V| + 1$ ומהיות דרגת הפיצול לכל היותר b הרי שהמילה שנוצרת מהעלים היא באורך לכל היותר $b^{|V|+1}$.

2. $|vy| > 0$ כי בחרנו את עץ הגזירה המינימלי ואם גם v וגם y היו ריקים, זה לא עץ גזירה מינימלי (הייתה לנו תת-גזירה $A \Rightarrow^* A$ כשיכלנו לדלג עליה, הביטו באיור לאינטואיציה).

3. $ux^i y v^i z \in L$ לכל $i \in \mathbb{N}_0$ כי אנחנו יודעים שניתן לגזור $S \Rightarrow^* uAz$ וגם $A \Rightarrow^* vAy$ ו- $A \Rightarrow^* x$ ולכן נוכל גם לבצע $A \Rightarrow^* v^i A y^i$ ולכן גם $ux^i y v^i z \in L$ ולכן גם $S \Rightarrow^* uv^i xy^i z$.

■

תרגול

דוגמאות

1. נביט בדקדוק הבא

$$A \rightarrow 0A1|B$$

$$B \rightarrow \#$$

במקרה כזה נוכל לגזור

$$A \rightarrow 0A1 \rightarrow 00A11 \rightarrow 00B11 \rightarrow 00\#11$$

וכמו שראינו בהרצאה, השפה היא כל המילים מהצורה $0^n \# 1^n$.

2. נמצא דקדוק לשפה $L = \{a^n b^{2n} : n \geq 0\}$. מספיק החוק $S \rightarrow aSbb \mid \epsilon$, כאשר כדי להגיע ל- $aabbbb$ נגזור

$$S \rightarrow aSbb \rightarrow aaSbbb \rightarrow aabbbb$$

3. $L = \{a^i b^j : j \geq i\}$. כאן נגדיר $S \rightarrow aSb \mid Sb \mid \epsilon$ ואז נוכל להוסיף כמה b -ים שנרצה מעבר ל- a -ים.

4. $L = \{a^i b^j c^j d^i : i, j \in \mathbb{N}_0\}$. כאן נגדיר

$$S \rightarrow aSd \mid T \mid \epsilon$$

$$T \rightarrow bTc \mid \epsilon$$

וזה יספיק כדי להגדיר לנו (המשתנה הנוסף כאן אינו הכרחי למעשה).

טענה CFL סגורה לאיחוד.

הוכחה: תהיינה $L_1, L_2 \in \text{CFL}$ ו- G_1, G_2 CFG-ים המתאימים להן. נוכיח כי $L_1 \cup L_2 \in \text{CFL}$. נניח בה"כ כי $V_1 \cap V_2 = \emptyset$ (שינוי שמות).

נגדיר

$$G = \langle V_1 \cup V_2 \cup \{S\}, \Sigma, R_1 \cup R_2 \{S \rightarrow S_1 \mid S_2\}, S \rangle$$

כאשר $S \notin V_1 \cup V_2$. קל לראות מכאן שמילה מתקבלת ע"י $S \rightarrow S_1 \rightarrow \dots$ כאשר היא יכולה להמשיך מ- S_1 . בגזירה המקורית שלה (ומ- S_2 בדומה). ■

טענה CFL סגורה לשרשור.

הוכחה: תהיינה $L_1, L_2 \in \text{CFL}$ ו- G_1, G_2 CFG-ים המתאימים להן. נוכיח כי $L_1 \cdot L_2 \in \text{CFL}$. נניח בה"כ כי $V_1 \cap V_2 = \emptyset$ (שינוי שמות).

נגדיר

$$G = \langle V_1 \cup V_2 \cup \{S\}, \Sigma, R_1 \cup R_2 \{S \rightarrow S_1 \cdot S_2\}, S \rangle$$

כאשר $S \notin V_1 \cup V_2$. קל לראות שכל השרשורים מתקבלים כאן, וכל מה שאינו שרשור לא מתקבל. ■

משפט (למת הניפוח ל-CFL) תהי $L \in \text{CFL}$ אזי קיים $p \geq 0$ קבוע הניפוח כך שלכל מילה $w \in L$ עם $|w| \geq p$, ניתן לכתוב $w = uvxyz$

כאשר מתקיים

$$1. |vxy| \leq p.$$

$$2. |vy| > 0.$$

$$3. \text{ לכל } i \in \mathbb{N}_0, \text{ מתקיים } uv^i xy^i z \in L.$$

הערה למה זו דומה מאוד ללמת הניפוח המקורית רק שעכשיו יש לנו שני סגמנטים שונים לנפח (ורק אחד מהם יכול להיות ריק), והחלק שהוא לכל היותר p לא חייב להיות רישא של המילה.

דוגמה $L = \{a^n b^n a^n : n \in \mathbb{N}\}$. נראה ש- $L \notin \text{CFL}$. נניח בשלילה ש- $L \in \text{CFL}$. יהי $p \geq 0$ קבוע ניפוח ונבחר $w = a^p b^p a^p$ ותהי $w = uvxyz$ חלוקה של w שמקיימת את התנאים.

$$\bullet \text{ } vxy \subseteq a^p \text{ אז עבור } i = 0, \text{ מכיל פחות } a\text{-ים מ-} vxy \text{ כי } |vy| > 0 \text{ ולכן } uv^i xy^i z \notin L \text{ ולכן } uv^i xy^i z \notin L \text{ סתירה.}$$

$$\bullet \text{ } vxy \subseteq b^p \text{ אז אותו הדבר כנ"ל עובד רק עם } b\text{-ים.}$$

$$\bullet \text{ } vxy \subseteq a^p \text{ אז כנ"ל.}$$

$\bullet \text{ } vxy \subseteq a^p b^p$ כאשר אנחנו מכילים לפחות a אחד או לפחות b אחד, אז בהכרח שבניפוח עם $i = 0$ נקטין את מספר ה- a ים או ה- b ים וכך נצא מהשפה סתירה.

$$\bullet \text{ } vxy \subseteq b^p a^p \text{ כאשר אנחנו מכילים לפחות } b \text{ אחד או לפחות } a \text{ אחד, אותו טיעון כנ"ל עובד.}$$

דוגמה ננסה להבין האם CFL סגורה לחיתוך. $L_1 = \{a^n b^n a^m : n, m \geq 0\}$ היא כן ב-CFL עם הדקדוק

$$S \rightarrow XA$$

$$X \rightarrow aXb | \epsilon$$

$$A \rightarrow Aa | \epsilon$$

וגם $L_2 = \{a^n b^m a^m : n, m \geq 0\}$ היא גם ב-CFL עם הדקדוק

$$S \rightarrow Ax$$

$$X \rightarrow bXa | \epsilon$$

$$A \rightarrow Aa | \epsilon$$

אבל $L_1 \cap L_2 = L$ כאשר L היא מהדוגמה הנ"ל וראינו שהיא לא ב-CFL לכן CFL לא סגורה לחיתוך.

הערה CFL בנוסף אינה סגורה להשלמה כי אחרת מדה-מורגן $\overline{L_1 \cap L_2} = \overline{L_1} \cup \overline{L_2}$ ואז CFL הייתה סגורה לחיתוך והיא לא.

דוגמה $L = \{ww : w \in \{a, b\}^*\}$. האם L ב-CFL? לא! נשתמש בלמת הניפוח ל-CFL.

תהי $p > 0$, נבחר את המילה $w = a^p b^p a^p b^p \in L$, ותהי חלוקה $w = uvxyz$ שמקיימת את התנאים.

• אם $xy \subseteq a^p$ או $xy \subseteq b^p$ במופע הראשון או $xy \subseteq a^p$ או $xy \subseteq b^p$ במופע השני, אז מהיות $|vy| > 0$, ניפוח ב- $i = 0$ נקבל שב- $uv^i xy^i z$ יש פחות a -ים (או b -ים) מאשר בצד השני ולכן יצאנו מהשפה סתירה.

• אם $xy \subseteq a^p b^p$ במופע הראשון או השני כאשר xy מכיל לפחות a אם v לא ריק ולפחות b אחד אם y לא ריק. אז כשנפח עם $i = 0$ נקבל פחות a -ים או b -ים מאשר בצד השני סתירה.

• אם $xy \subseteq b^p a^p$ אז v לא ריק ואז הוא מכיל b או y לא ריק והוא מכיל a ואז נפח עם $i = 0$ ונקבל שאחד הצדדים קטן מהאחר קטן מהאחר ונצא מהשפה סתירה.

נביט ב- \bar{L} , זו השפה $\{uv : |u| = |v|, u \neq v\} \cup \{w : |w| \equiv 1 \pmod{2}\}$ (כל המילים האי זוגיות או זוגיות עם חצאים שונים). נוכיח כי $\bar{L} \in \text{CFL}$. L_2 רגולרית (אוטומט שבודק זוגיות של אורך המילה) ולכן ב-CFL. נותר להוכיח כי $L_1 \in \text{CFL}$ (למרות שהראנו כבר בהרצאה שהיא רגולרית) ונסיים מסגירות לאיחוד.

תהי $w \in L_1$, $|w| = 2n$ עבור $n \geq 0$ וקיים אינדקס $i \in [n]$ כך ש- $w_i = a$ ו- $w_{n+i} = b$ (בה"כ).

לכן ניתן לכתוב $w = xaybz$ כאשר $|x| = i - 1, |y| = n - 1, |z| = n - i$. נגדיר את L_1 מחדש בתור

$$\{xaybz, xbyaz : x, y, z \in \{a, b\}^*, |y| = |x| + |z|\}$$

נמצא דקדוק ל- L_1 .

$$S \rightarrow AB|BA$$

$$A \rightarrow aAa|aAb|bAa|bAb|a$$

$$B \rightarrow aBa|aBb|bBa|bBb|b$$

מזהה את L_1 כאשר הרעיון כאן הוא שהגזירה מ- S מחלקת אותנו להאם יש לנו a קודם ואז b או b קודם ואז a , וכדי לשמור על דרישת האורך, בכל פעם שנוסיף טרמינל ל- x או z נוסיף גם אות ל- y כדי לשמור על שוויון בין האגפים.