הרצאה 5

1 למת הניפוח לשפות רגולריות

הגדרה: קבוצת מצבים $C \subseteq Q$ תקרא **מעגל** אם: $\exists q \in C$ כך ש: $\hat{\delta}(q,w) = q$: $\exists w \in \Sigma^+$, (כלומר, אם קיים מצב q ב-C שעבורו קיימת מילה לא ריקה שאם נתחיל מ-q ונקרא את המילה, נחזור ל-q). ובנוסף, קריאת כל אות במילה משאיר את האוטומט במצב מ-C. פורמלית:

$$\{\forall u, v \neq w \; s.t \; uv = w : \; \hat{\delta}(q,u) \in C\}$$

1.1 בכל אס"ד קיים מעגל

הוכחה: אם קיימת ב-L מילה הארוכה ביותר, אזי קיים מצב בור עבור כל המילים שארוכות ממנה. בור הוא מעגל עבור כל מילה.

אם אין מילה ארוכה ביותר, אזי $L(A)$ אינסופית. היות ו- $|Q|$ סופי (כי זה **אס**"ד), בהכרח קיים מעגל עבור מילים שארוכות מ-$|Q|$ (שובך היונים): נב"ש שאין מעגלים. נבחר מילה שארוכה יותר מ-$|Q|$. אין מעגלים, כלומר בכל צעד הגענו למצב חדש. המצבים ייגמרו לפני שנסיים את המילה. הצעד הבא חייב להיות למצב שכבר היינו בו.

שפה היא רגולרית אמ"מ קיים אס"ד המקבל אותה. כל שפה סופית היא רגולרית. קיימות שפות לא רגולריות – ראינו הוכחה ישירה לשפה ספציפית. בהוכחה שראינו, השתמשנו בסופיות מספר מצבי האוטומט כדי לטעון שיש סוג של מחזוריות באופן הפעולה שלו. ננסה להכליל את הטיעון כדי לקבל תנאי הכרחי להיותה של שפה רגולרית:

1.2 למת הניפוח

תהי L שפה רגולרית. אזי, קיים $n \in \mathbb{Z}^+$ שעבורו לכל $z \in L$ המקיימת $n \leq |z|$, קיים פירוק $z = uvw$ המקיים:

א. $|uv| \leq n$

ב. $1 \leq |v|$

ג. $\forall i \in \mathbb{Z}^+ \cup \{0\} : uv^i w \in L$

כלומר: יש אורך מסוים, שכל מילה שארוכה ממנו מקיימת את התנאים.

למת הניפוח היא תנאי **הכרחי אך לא מספיק** עבור רגולריות. נראה בהמשך שפות שמקיימות את הלמה, אבל אינן רגולריות. כלומר, עיקר השימוש של הלמה יהיה כדי להפריך רגולריות של שפה.

1.3 הוכחת הלמה

תהי L שפה רגולרית, ויהי $A = (Q, \Sigma, q_0, \delta, F)$ אס"ד המקבל אותה. יהי $n = |Q|$.

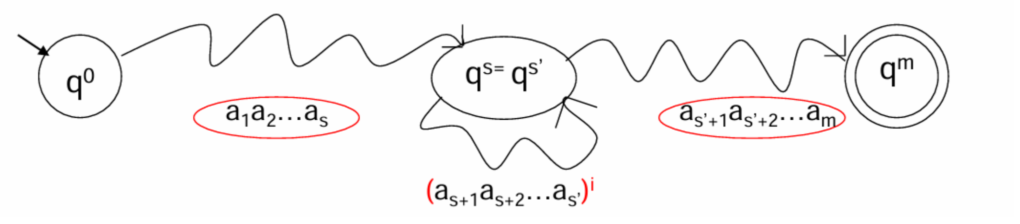
תהי $z = a_1 a_2 \cdots a_m$ מילה כלשהי ב-L, כך ש $m \geq n$.

לכל $0 \leq i \leq m$ נסמן $q^i = \hat{\delta}(q_0, a_1 a_2 \cdots a_i)$, כאשר $q^0 = q_0$. (המצב אחרי קריאת i תווים. $m+1$ יונים).

כיוון ש- $n = |Q|$ (שובכים), קיימים זוג אינדקסים s, s' המקיימים: $0 \leq s < s' \leq m$ כך ש: $q^s = q^{s'}$.

שימו לב ש $s \leq n$, כי כבר ברישא $a_1 a_2 \cdots a_n$ קיימת חזרה על מצב פעמיים, (כי מילה באורך n צריכה $n+1$ מצבים, כי צריך מצב גם עבור המילה הריקה). ובחרנו את s, s' לפני החזרה הראשונה.

נתבונן בחישוב האוטומט על המילה $z = a_1 a_2 \cdots a_m$:



מכיוון ש $z \in L$, נקבל ש $q^m \in F$. ונובע גם ש $a_1a_2\cdots a_sa_{s+1}\cdots a_m \in L$.

נוכיח באינדוקציה ש $a_1a_2\cdots a_s(a_{s+1}\cdots a_{s'})^i a_{s'+1}\cdots a_m \in L$, לכל $0 \le i$. נרשום:

$$z = \underbrace{a_1a_2\cdots a_s}_{u}\underbrace{a_{s+1}\cdots a_{s'}}_{v}\underbrace{a_{s'+1}\cdots a_m}_{w}$$

נראה כי שלושת תנאי הלמה מתקיימים:

א. $|uv| = s' \le n$ (לפי בחירת s, s')
ב. $1 \le |v|$ (כי $s < s'$)

נותר רק להוכיח את:

ג. $\forall i \in \mathbb{Z}^+ \cup \{0\} : uv^iw \in L$

עלינו להוכיח כי בפירוק הנ"ל, $\forall i \ge 0 : uv^iw \in L$, כאשר אנחנו יודעים ש $\hat{\delta}(q^s, v) = q^s$.

נוכיח ראשית כי לכל i מתקיים $\hat{\delta}(q^s, v^i) = q^s$, באינדוקציה על i.

בסיס: $i = 0$, $v^0 = \epsilon$, $\delta(q^s, \epsilon) = q^s$.

צעד: נניח ש $\hat{\delta}(q^s, v^i) = q^s$. מכאן נקבל: א – מהנ"א.

$$\hat{\delta}(q^s, v^{i+1}) = \hat{\delta}(q^s, v^iv) = \hat{\delta}(\hat{\delta}(q^s, v^i), v) \overset{\text{א}}{=} \hat{\delta}(q^s, v) = q^s$$

אז הוכחנו ש $\forall i \ge 0 : uv^iw \in L$.

כמו כן אנחנו יודעים ש $\hat{\delta}(q^s, w) = q^m \in F$, וכן $\hat{\delta}(q^0, u) = q^s$. נרשום: א – מהאינדוקציה שהוכחנו.

$$\hat{\delta}(q^0, uv^iw) = \hat{\delta}(\hat{\delta}(q^0, uv^i), w) = \hat{\delta}(\hat{\delta}(\hat{\delta}(q^0, u), v^i), w) = \hat{\delta}(\hat{\delta}(q^s, v^i), w) \overset{\text{א}}{=} \hat{\delta}(q^s, w) = q^m \in F$$

1.4 למת הניפוח בשפות סופיות

שימו לב, שלמת הניפוח לא מוגדרת רק עבור שפות אינסופיות. איך היא מתקיימת בשפות סופיות?

יהי t אורך המילה הארוכה ביותר בשפה. נבחר $n = t + 1$, ואכן מתקיים שכל מילה שאורכה לפחות n ניתנת לניפוח – אין מילים כאלה, אז הטענה מתקיימת באופן ריק.

1.5 הוכחת אי – רגולריות ע"י הלמה

כדי להוכיח ששפה אי – רגולרית:

נניח בשלילה שהיא רגולרית. ניקח את ה-n שקיומו מובטח בלמה. נבחר מילה z באורך $|z| \ge n$. נקבל פירוק כלשהו $z = uvw$. ("האויב" בוחר את הפירוק). מוצאים $i \ne 1$ שעבורו $uv^iw \notin L$, וכך מגיעים לסתירה.

למה תמיד $i \ne 1$? כי עבור $i = 1$, זה פשוט z. וברור שהיא שייכת לשפה.

2 דוגמאות לשימוש בלמת הניפוח

2.1 דוגמה 1

צ"ל שהשפה $L = \{x \in \{a,b\}^* : \#_a(x) = \#_b(x)\}$ אינה רגולרית.

נב"ש שהיא כן רגולרית. יהי n הקבוע שקיומו מובטח בלמת הניפוח. נבחר $z = a^n b^n$. בבירור $z \in L$, וגם $|z| \geq n$. יהי uvw פירוק של z כפי שמובטח בלמה. נבחר $i = 0$ ונגיע לסתירה: כיוון שמתקיים $|uv| \leq n$ וגם $1 \leq |v|$, נקבל:

$$z_0 = uv^0w = uw = a^{n-|v|}b^n \notin L$$

בסתירה ללמת הניפוח. כלומר, L אינה רגולרית.

שימו לב ש $a^{n-|v|}b^n \notin L$ בגלל שחרגנו מההגדרה על מספר ה-a,b ולא בגלל שחרגנו מהמבנה של $a^k b^k$.

2.2 דוגמה 2

צ"ל שהשפה $L = \{xx : x \in \{a,b\}^*\}$ אינה רגולרית.

נב"ש שהיא כן רגולרית. יהי n הקבוע שקיומו מובטח בלמת הניפוח. נבחר $z = a^n b a^n b$. יהיה קל לטפל במילה הזאת, כי ה-n התווים הראשונים זהים.

בבירור $z \in L$, וגם $|z| \geq n$. יהי uvw פירוק של z כפי שמובטח בלמה. בגלל ש ה-n התווים הראשונים זהים, כיוון שמתקיים $|uv| \leq n$, uv כולה מורכבת מ-a. וגם נתון ש $1 \leq |v|$.

נבחר $i = 0$ ונגיע לסתירה: נקבל:

$$z_0 = uv^0w = uw = a^{n-|v|}ba^nb \notin L$$

בסתירה ללמת הניפוח. כלומר, L אינה רגולרית.

2.3 דוגמה 3

צ"ל שהשפה $L = \{a^{k^2} : k \in \mathbb{N}\}$ אינה רגולרית.

נב"ש שהיא כן רגולרית. יהי n הקבוע שקיומו מובטח בלמת הניפוח. נבחר $z = a^{n^2}$. בבירור $z \in L$, וגם $|z| \geq n$. יהי uvw פירוק של z כפי שמובטח בלמה. נסמן $t = |v|$ (עבור $1 \leq t \leq n$ כלשהו).

אנחנו רוצים ש $uv^iw \notin L$. מתקיים: $uv^iw = a^{n^2+(i-1)|v|}$. נרצה למצוא i כך ש $n^2 + (i-1)|v|$ הוא לא ריבוע של אף מספר שלם. ננסה למצוא מספר בין הריבוע של n, לריבוע הבא. כלומר משהו שקטן מ: $(n+1)^2 = n^2 + 2n + 1$.

אם נוסיף ל-n^2 מספר שהוא בין 1 ל- $2n+1$, זה יעבוד. נזכר ש $1 \leq |v| \leq n$, אז נביא למצב שמוסיפים את $|v|$.

נבחר $i = 2$ ונגיע לסתירה:

$$z_2 = uv^2w = uw = a^{n^2+t} \notin L$$

(כי $n^2 < n^2 + 1 \leq n^2 + t \leq n^2 + n < n^2 + 2n + 1 = (n+1)^2$, כלומר העלנו את a "בחזקת" מספר שהוא לא k^2 לאף k).

בסתירה ללמת הניפוח. כלומר, L אינה רגולרית.

3 שפה לא רגולרית הניתנת לניפוח

כאמור, למת הניפוח היא תנאי הכרחי אך לא מספיק לרגולריות. נוכיח את זה: נראה שפות לא רגולריות שכן ניתנות לניפוח.

לדוגמה: $L = \{a\}^* \cup \{b^j a^{k^2} : 1 \le j,k\}$.

3.1 קיום למת הניפוח

נראה שלמת הניפוח מתקיימת, עם $n = 1$. עבור כל $z \in L$ כך ש $|z| \ge 1$, נגדיר פירוק $z = uvw$ כך ש $u = \epsilon, |v| = 1$.

אם $z \in \{a\}^*$, אז לכל i גם $z_i = uv^iw \in \{a\}^* \subseteq L$.

אם z מהצורה $b^j a^{k^2}$ כאשר $j > 0$, אזי לכל i גם z_i מהצורה הזו. כלומר, למת הניפוח מתקיימת.

3.2 אי – רגולריות

נראה שהשפה לא רגולרית: נב"ש שהיא כן רגולרית. ניזכר שהשפה $L_1 = \{a\}^*$ רגולרית. אזי, מסגירות להפרש נקבל:

$L' = L \setminus L_1 = \{b^j a^{k^2} : 1 \le j,k\}$ רגולרית. מסגירות להיפוך נובע שגם $(L')^R = \{a^{k^2} b^j : 1 \le j,k\}$ רגולרית.

אבל ע"י אותה הוכחה מ-2.3, מתקבל כי $(L')^R$ לא ניתנת לניפוח. סתירה לכך שהיא רגולרית.

4 תרגילים ממבחנים

הוכיחו או הפריכו: השפות הבאות רגולריות:

א: $L_1 = \{a^i b^j : i \le j\}$. אינטואיטיבית, זה אוטומט שדורש לספור. אין אוטומט כזה. נוכיח שלא רגולרית:

נב"ש שכן רגולרית, כלומר למת הניפוח מתקיימת. יהי n הקבוע המובטח מהלמה. תהי $z = a^n b^n$. מתקיים $|z| = 2n \ge n$ ולכן קיים פירוק $z = uvw$ המקיים $|v| \ge 1, |uv| \le n$. נשים לב ש uv מורכבת רק מ-a.

עבור $i = 2$ נקבל שהמילה המנופחת היא: $uv^2w = a^{n+(i-1)|v|} b^n = a^{n+2|v|} b^n$, לא בשפה.

ב: $L_2 = \{a^i b^j : i \ne j\}$

נב"ש שהיא רגולרית ויהי n הקבוע המובטח בלמה. תהי $z = a^n b^{n+n!}$. $z \in L$ וגם $|z| \ge n$. ולכן קיים פירוק $z = uvw$ כך ש:

$$|uv| \le n, \qquad |v| \ge 1, \qquad \forall i \in \mathbb{N}: uv^iw \in L$$

לכל i נקבל:

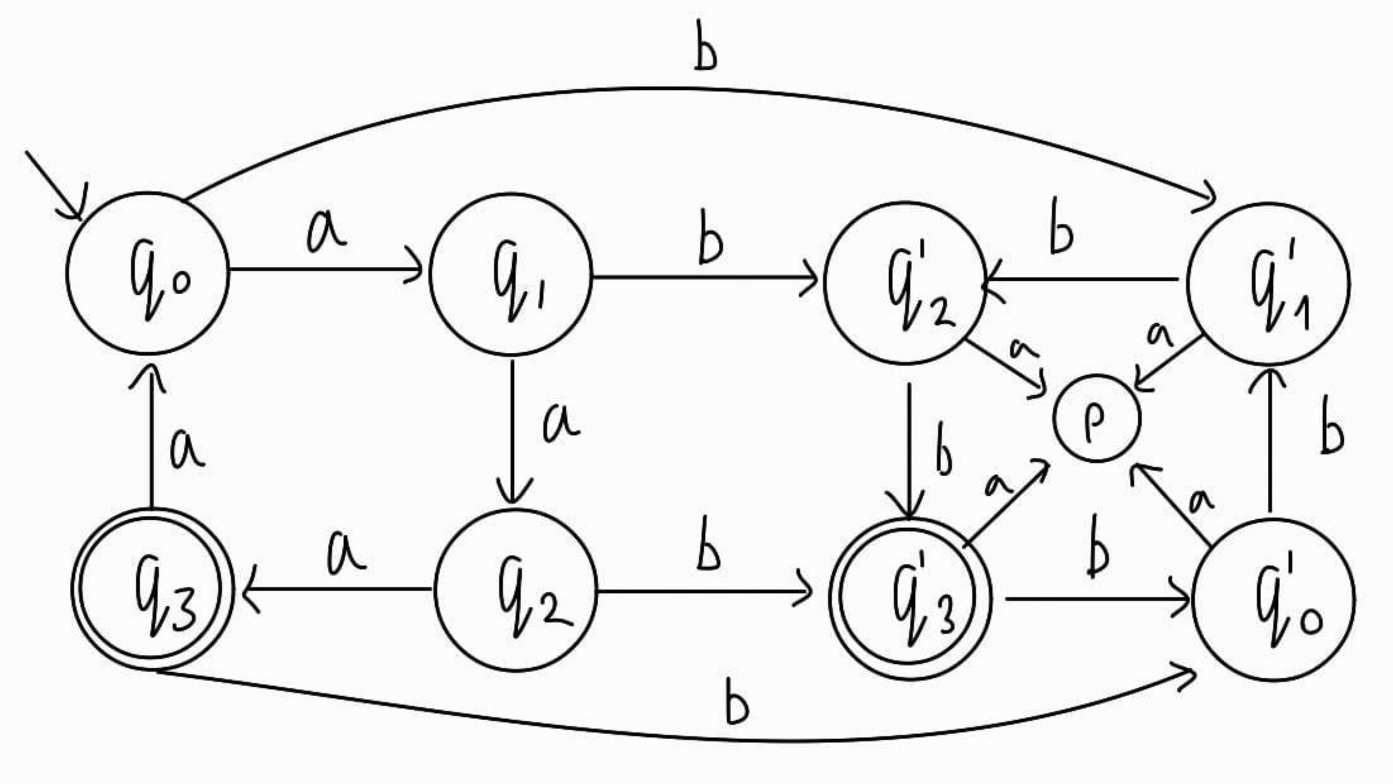
$$uv^iw = a^n a^{(i-1)\cdot|v|} b^{n+n!} = a^{n+(i-1)|v|} b^{n+n!}$$

מתי $n + (i-1)|v| = n + n!$?

$$(i-1)|v| = n! \to i|v| - |v| = n! \to i|v| = n! + |v| \to i = \frac{n!}{|v|} + 1$$

ג: $L = \{a^i b^j : i + j \equiv 3 (mod\ 4)\}$

בתור כלל אצבע, שפה עם מודולו תהיה רגולרית. נבנה אוטומט ונהפוך אותו לב"ר:



$$(q_0 \to q_0) = a^4, (q_0 \to q_3) = a^3, (q_3 \to q_3) = \epsilon, (q_0 \to q_3') = b^3 + ab^2 + a^2b + a^3b^4, (q_3' \to q_3') = b^4$$

$$(a^4)^*[a^3 + (b^3 + ab^2 + a^2b + a^3b^4)(b^4)^*] = (a^4)^*[a^3 + (b^3 + ab^2 + a^2b)(b^4)^*]$$

אפשר גם ישירות, בלי אוטומט:

$$(a^4)^*(a^3 + a^2b + ab^2 + b^3)(b^4)^*$$

בהתחלה ובסוף, כפולות של 4 לא משפיע על מודולו. ובאמצע, כל הדרכים להגיע לסכום שהוא 3 מוד 4.

ד: $L_4 = \{a^ib^jc^k : k^2 < j < 10i \wedge 0 \le i < k\}$

לכאורה, לא רגולרי כי צריך לספור. אבל נשים לב להגבלות:

אם $k^2 < j < 10i$ וגם $i < k$, זה אומר ש $k^2 < 10k$, כלומר $k < 10$. זה מגביל את i ולכן גם את j. $i < 9$, $j < 90$.

נכתוב מחדש את ההגדרה:

$$L_4 = \{a^ib^jc^k : k < 10, i < 9, j < 90\}$$

השפה סופית ולכן רגולרית.

ה: $L_5 = \{a^ib^jc^k : \min(i,j) \le k\}$

נב"ש שהיא רגולרית ויהי n הקבוע המובטח בלמה. אנחנו רוצים מילה שאם ננפח אותה, התנאי לא יתקיים. כלומר שהקטן מבין מספר ה-a,b יהיה גדול ממש ממספר ה-c.

תהי $z = a^nb^{2n}c^n$. $z \in L$ וגם $|z| \ge n$. ולכן קיים פירוק $z = uvw$ כך ש:

$$|uv| \le n, \qquad |v| \ge 1, \qquad \forall i \in \mathbb{N}: uv^iw \in L$$

כל ניפוח רק יגדיל את מספר ה-a, לדוגמה $i = 2$:

$$uv^2w = a^{n+|v|}b^{2n}c^n \notin L_5$$

ו: $L_6 = \{w \in \{0,1\}^* : \#_0(w) \equiv 1(mod\ 3) \wedge \#_1(w) \equiv 2(mod\ 3)\}$

אוטומט לכל אחד, ומכפלה.

תרגיל: הוכיחו/הפריכו:

א. אם L שפה שמקיימת את למת הניפוח עבור $n = 20$, אז היא מקיימת את למת הניפוח עבור $n = 17$.
ב. אם L שפה שמקיימת את למת הניפוח עבור $n = 17$, אז היא מקיימת את למת הניפוח עבור $n = 20$.

אינטואיציה:

אם כל מילה שארוכה מ20 מקיימת תנאי כלשהו, אז עדיין לא בהכרח כל מילה שארוכה מ17 מקיימת את התנאי. אז נמצא מילה בין 17-19 שלא מקיימת, וזה יפריך.

אם כל מילה שארוכה מ17 מקיימת תנאי, אז ברור שכל מילה שארוכה מ20 מקיימת.

פורמלית:

א. נפריך: נקח את השפה $L = \{w \in \Sigma^* : |w| \equiv 18\ (mod\ 20)\}$. לכל מילה ארוכה מ20, הפירוק: $u = \epsilon, |v| = 20$ מקיים את התנאי. אבל מילה באורך 17 לא תקיים, כי אין מעגל באורך 17.

תרגיל נוסף: תהי $L = \{w \in \{0,1\}^* : \#_0(w) \equiv \{2,3\}(mod\ 5)\}$

הראו כי השפה מקיימת את למת הניפוח. (בפרט, קבעו את ה-n המינימלי שעבורו הלמה מתקיימת).

נקבע $n = 5$. לכל $z \in L$ כך ש $n \leq |z|$ נקבע את הפירוק: $z = uvw$ כך שלכל $i \in \mathbb{N}$, $uv^iw \in L$.

אם 5 התווים הראשונים הם 0, נקבע $u = \epsilon, v = 0^5$. ואז כל ניפוח ישאיר את מספר האפסים שקול במוד 5.

אם ב5 התווים הראשונים יש 1, נקבע: u הוא מספר האפסים עד ה1 הראשון. $v = 1$. ואז כל ניפוח לא משפיע על מספר האפסים.