פרק 4
שפות רגולריות

4.1 אוטומט המכפלה

הגדרה 4.1 אטומט המכפלה

יהיו

$$M_1 = (Q_1, \Sigma, \delta_1, q_{0,1}, F_1)$$
$$M_2 = (Q_2, \Sigma, \delta_2, q_{0,2}, F_2)$$

שני אוטומטים דטרמיניסטיים המקבלים את השפות L_1, L_2 בהאמה. **האוטומט המכפלה** תסומן $M_1 \times M_2$ ויוגר

$$M_1 \times M_2 = (Q_1 \times Q_2, \Sigma, \delta, (q_{0,1}, q_{0,2}), F_1 \times F_2)$$

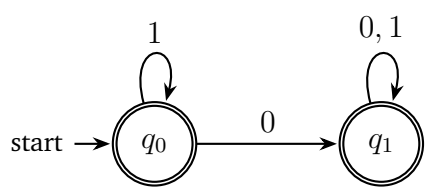
כאשר לכל $q_1 \in Q_1$ ולכל $q_2 \in Q_2$ ולכל $\sigma \in \Sigma$:

$$\delta\left((q_1, q_2), \sigma\right) = \left(\delta_1(q_1, \sigma)\ ,\ \delta_2(q_2, \sigma)\right)\ .$$

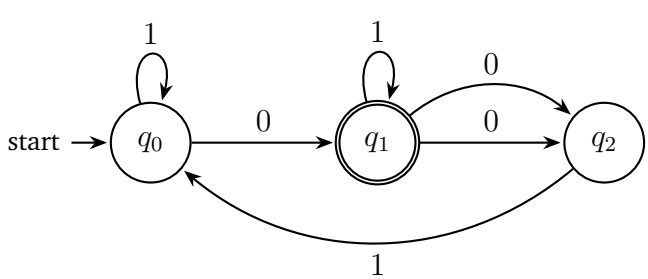
דוגמה 4.1

בתרשים נתונים שני אוטומטים דטרמיניסטיים M_1, M_2.

M_1

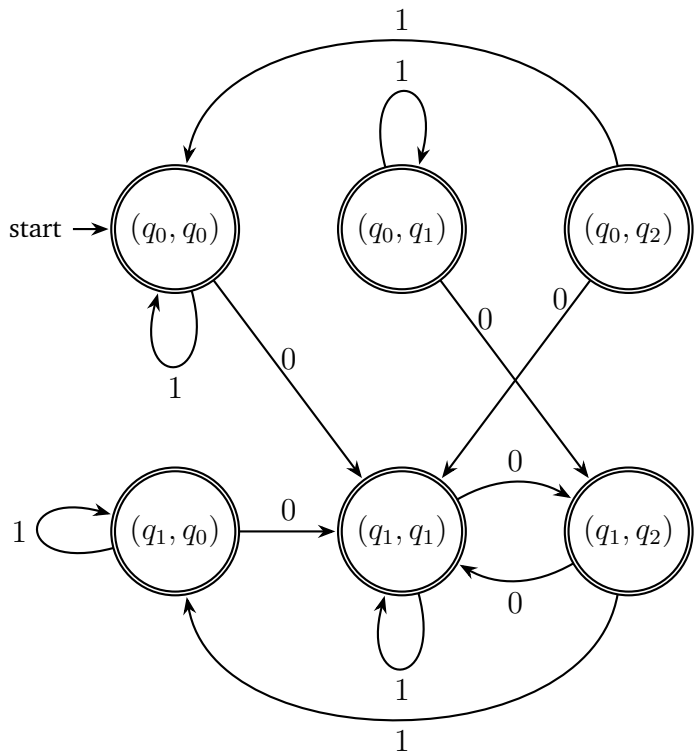


M_2



ציירו את הרשים של האוטומט המכפלה.

פתרון:



4.2 סגירות

הגדרה 4.2

יהי Σ אלפבית נתון (קבוצה סופית של תווים) ותהיינה L_1, L_2 שפות מעל האלפבית הזה. נמנה מספר פעולות שבאמצעותן ניתן לייצר שפות חדשות משפות נתונות כזכור:

- **האיחוד** של השפות L_1, L_2 מוגדר

$$L_1 \cup L_2 = \{w \mid w \in L_1 \ \vee \ w \in L_2\} \ .$$

- **החיתוך** של השפות L_1, L_2 מוגדר

$$L_1 \cap L_2 = \{w \mid w \in L_1 \ \wedge \ w \in L_2\} \ .$$

- **השרשור** של השפות L_1, L_2 מוגדר

$$L_1 L_2 = \{u\mathrm{v} \mid u \in L_1 \ \wedge \ \mathrm{v} \in L_2\} \ .$$

- **החזקה** של השפה L מוגדרת:

$$\begin{aligned} L^0 &= \{\varepsilon\} \ , \\ L^1 &= L \ , \\ L^2 &= LL \ , \\ L^{n+1} &= L^n L \ . \end{aligned}$$

- **הסגור קליין** מוגדר
$$L^* = \bigcup_{n=0}^{\infty} L^n \ .$$
- **הסגור קליין פלוס** מוגדר
$$L^+ = \bigcup_{n=1}^{\infty} L^n \ .$$

משפט 4.1 סגירות שפות רגולריות תחת חיתוך

אם L_1, L_2 שפות רגולראיות אז $L_1 \cap L_2$ היא שפה רגולרית.

הוכחה:

- L_1 שפה רגולרית $\Longleftarrow$ $\exists$ אוטומט $A_1 = (Q_1, \Sigma_1, \delta_1, q_0, F_1)$ שמקבלת אותה.
- L_2 שפה רגולרית $\Longleftarrow$ $\exists$ אוטומט $A_2 = (Q_2, \Sigma_2, \delta_2, p_0, F_2)$ שמקבלת אותה.
- קיים אוטומט דטרמיניסטי A המקבל את השפה $L_1 \cap L_2$, והוא האוטומא המכפלה:
$$A = A_1 \times A_2 = (Q_1 \times Q_2, \Sigma, \delta, (q_0, p_0), F_1 \times F_2) \ .$$

כאשר פונקציית המעברים δ מוגדרת כך (סימולציה במקביל):
$$\delta((q,p),\sigma) = (\delta_1(q,\sigma), \delta_2(p,\sigma)) \ .$$

הוכחת נכונות.

כעת נוכיח כי $A_1 \times A_2$ מקבל את השפה $L_1 \cap L_2$ באופן הבא.
נשים לב שעל פי הגדרת פונקציית המעברים, לכל מילה w, מסלול החישוב משמר את המעברים של כל אוטומט בנפרד:
$$\delta^*((q_0,p_0),w) = (\delta_1^*(q_0,w), \delta_2^*(p_0,w))$$

כיוון 1:

אם $w \in L_1 \cap L_2$

$\Longleftarrow$ $w \in L_1$ וגם $w \in L_2$.

$\Longleftarrow$ A_1 מקבל את w וגם A_2 מקבל את w.

$\Longleftarrow$ הרצה של A_1 על w מסתיימת במצב $q_f \in F_1$,
וגם הרצה של A_2 על w שמסיימת במצב $p_f \in F_2$.

$\Longleftarrow$ על פי תכונת δ^*, הרצה של A על w מסתיימת במצב $(q_f, p_f) \in F_1 \times F_2$ אשר שייך לקבוצת המצבים המקבלים $F_1 \times F_2$.

$\Longleftarrow$ A מקבל את w.

כיוון 2:

אם $w \notin L_1 \cap L_2$

$\Longleftarrow$ $w \notin L_1$ או $w \notin L_2$. נניח בלי הגבלת הכלליות ש-$w \notin L_1$.

$\Longleftarrow$ A_1 לא מקבל את w.

$\Longleftarrow$ הרצה של A_1 על w מסתיימת במצב $q \notin F_1$.

$\Longleftarrow$ הרצה של A על w מסתיימת במצב $(q,p) \notin F_1 \times F_2$.

$\Longleftarrow$ הרצה של A על w מסתיימת במצב שאינו מצב מקבל.

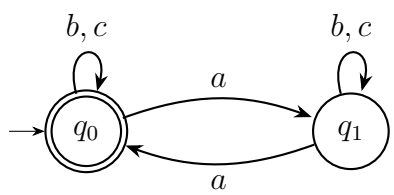
$\Longleftarrow$ A לא מקבל את w.

■

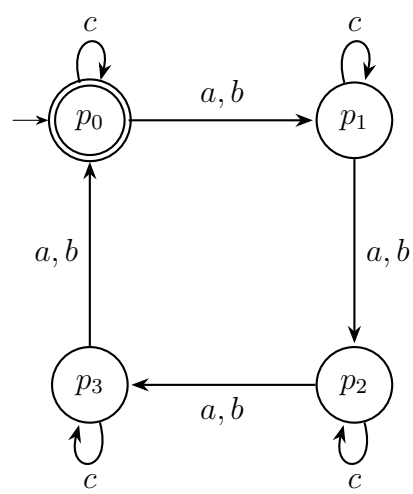
דוגמה 4.2

נתונים האוטומטים הבאים:

M_1



M_2



א) מהן השפות $L(M_1)$ ו- $L(M_2)$?

ב) בנו אוטומט המקבל את השפה $L(M_1) \cap L(M_2)$.

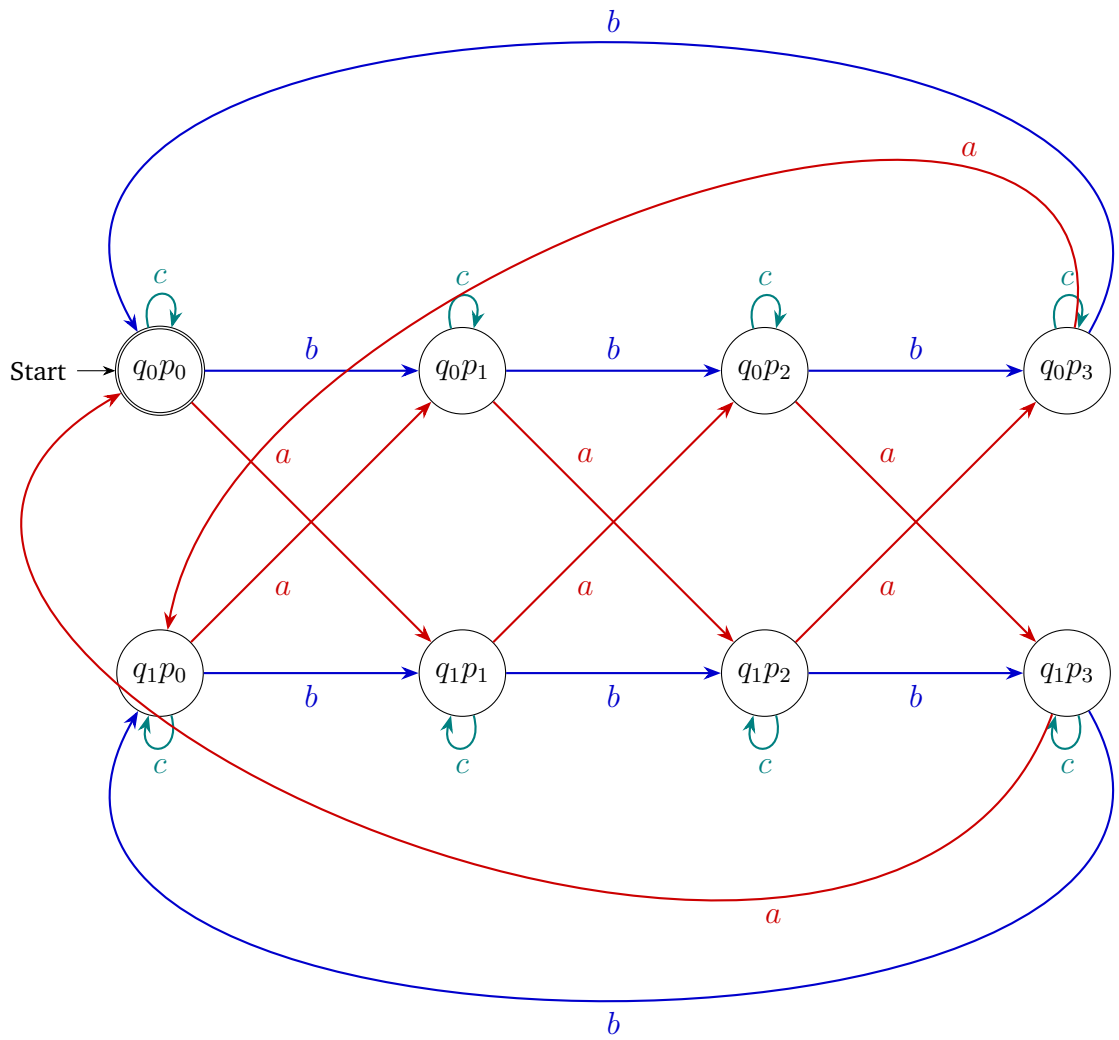
פתרון:

סעיף א)

$$L(M_1) = \left\{ w \in \{a,b,c\}^* \;\middle|\; \#a_w \bmod 2 = 0 \right\} \ ,$$
$$L(M_2) = \left\{ w \in \{a,b,c\}^* \;\middle|\; (\#a_w + \#b_w) \bmod 4 = 0 \right\} \ .$$

סעיף ב)

$$L(M_1) \cap L(M_2) = \left\{ w \in \{a,b,c\}^* \;\middle|\; (\#a_w \bmod 2 = 0) \wedge ((\#a_w + \#b_w) \bmod 4 = 0) \right\} \ .$$



משפט 4.2 סגירות שפות רגולריות תחת איחוד

אם L_1, L_2 שפות רגולראיות אז $L_1 \cup L_2$ היא שפה רגולרית.

הוכחה:

- L_1 שפה רגולרית $\Longleftarrow$ $\exists$ אוטומט $A_1 = (Q_1, \Sigma_1, \delta_1, q_0, F_1)$ שמקבלת אותה.
- L_2 שפה רגולרית $\Longleftarrow$ $\exists$ אוטומט דטרמיניסטי $A_2 = (Q_2, \Sigma_2, \delta_2, p_0, F_2)$ שמקבלת אותה.
- קיים אוטומט A המקבל את השפה $L_1 \cup L_2$, והוא האוטומא המכפלה הבא:

$$A = (Q_1 \times Q_2, \Sigma, \delta, (q_0, p_0), (F_1 \times Q_1) \cup (Q_1 \times F_2)) \ .$$

כאשר הפונקציית המעברים δ מוגדרת כמו באוטומט מכפלה:

$$\delta((q,p),\sigma) = (\delta_1(q,\sigma), \delta_2(p,\sigma)) \ .$$

הוכחת נכונות.

כעת נוכיח כי
$$A = (Q_1 \times Q_2, \Sigma, \delta, (q_0, p_0), (F_1 \times Q_1) \cup (Q_1 \times F_2))$$
מקבל את השפה $L_1 \cup L_2$ באופן הבא.

כיוון 1:

אם $w \in L_1 \cup L_2$

$\Longleftarrow$ $w \in L_1$ או $w \in L_2$.

$\Longleftarrow$ A_1 מקבל את w או A_2 מקבל את w.

$\Longleftarrow$ הרצה של A_1 על w מסתיימת במצב $q_f \in F_1$,
או הרצה של A_2 על w שמסתיימת במצב $p_f \in F_2$.

$\Longleftarrow$ על פי תכונת δ^*, הרצה של A על w מסתיימת במצב (q,p) כך ש- $q \in F_1$ או $p \in F_2$.

$\Longleftarrow$ הרצה של A על w מסתיימת במצב (q,p) כך ש- $(q,p) \in (F_1 \times Q_2) \cup (Q_1 \times F_2)$, שהוא קבוצת המצבים המקבלים של האוטומט.

$\Longleftarrow$ A מקבל את w.

כיוון 2:

אם $w \notin L_1 \cup L_2$

$\Longleftarrow$ $w \notin L_1$ **וגם** $w \notin L_2$.

$\Longleftarrow$ A_1 לא מקבל את w וגם A_2 לא מקבל את w.

$\Longleftarrow$ הרצה של A_1 על w מסתיימת במצב $q \notin F_1$ וגם הרצה של A_2 על w מסתיימת במצב $p \notin F_1$.

$\Longleftarrow$ הרצה של A על w מסתיימת במצב $(q,p) \notin (F_1 \times Q_2) \cup (Q_1 \times F_2)$.

$\Longleftarrow$ הרצה של A על w מסתיימת במצב שאינו מצב מקבל.

$\Longleftarrow$ A לא מקבל את w.

■

משפט 4.3 סגירות שפות רגולריות תחת הפרש

אם L_1, L_2 שפות רגולראיות אז $L_1 \backslash L_2$ היא שפה רגולרית.

הוכחה:

- L_1 שפה רגולרית $\Longleftarrow$ $\exists$ אוטומט $A_1 = (Q_1, \Sigma_1, \delta_1, q_0, F_1)$ שמקבלת אותה.
- L_2 שפה רגולרית $\Longleftarrow$ $\exists$ אוטומט $A_2 = (Q_2, \Sigma_2, \delta_2, p_0, F_2)$ שמקבלת אותה.

- קיים אוטומט דטרמיניסטי A המקבל את השפה $L_1 \cup L_2$, והוא האוטומא המכפלה הבא:

$$A = (Q_1 \times Q_2, \Sigma, \delta, (q_0, p_0), (F_1 \times Q_2 \backslash F_2)) \ .$$

כאשר הפונקציית המעברים δ מוגדרת כמו באוטומט מכפלה:

$$\delta((q,p),\sigma) = (\delta_1(q,\sigma), \delta_2(p,\sigma)) \ .$$

הוכחת נכונות.

כעת נוכיח כי

$$A = (Q_1 \times Q_2, \Sigma, \delta, (q_0, p_0), (F_1 \times Q_2 \backslash F_2))$$

מקבל את השפה $L_1 \backslash L_2$ באופן הבא.

כיוון 1:

אם $w \in L_1 \backslash L_2$

$\Longleftarrow$ $w \in L_1$ וגם $w \notin L_2$.

$\Longleftarrow$ A_1 מקבל את w וגם A_2 לא מקבל את w.

$\Longleftarrow$ הרצה של A_1 על w מסתיימת במצב $q_f \in F_1$,
וגם הרצה של A_2 על w לא מסתיימת במצב $p_f \in F_2$.

$\Longleftarrow$ על פי תכונת δ^*, הרצה של A על w מסתיימת במצב (q,p) כך ש- $q \in F_1$ או $p \notin F_2$.

$\Longleftarrow$ הרצה של A על w מסתיימת במצב (q,p) כך ש- $(q,p) \in (F_1 \times Q_2 \backslash F_2)$, שהוא קבוצת המצבים המקבלים של האוטומט.

$\Longleftarrow$ A מקבל את w.

כיוון 2:

אם $w \notin L_1 \backslash L_2$.

$\Longleftarrow$ $w \in \bar{L}_1 \cup L_2$.

$\Longleftarrow$ $w \notin L_1$ או $w \in L_2$.

$\Longleftarrow$ A_1 לא מקבל את w או A_2 מקבל את w.

$\Longleftarrow$ הרצה של A_1 על w מסתיימת במצב $q_f \notin F_1$,
וגם הרצה של A_2 על w שמסתיימת במצב $p_f \in F_2$.

$\Longleftarrow$ על פי תכונת δ^*, הרצה של A על w מסתיימת במצב (q,p) כך ש- $q \notin F_1$ או $p \in F_2$.

$\Longleftarrow$ הרצה של A על w מסתיימת במצב (q,p) כך ש- $(q,p) \in (Q_1 \backslash F_1 \times F_2)$, שהוא לא קבוצת המצבים המקבלים של האוטומט A.

$\Longleftarrow$ A לא מקבל את w.

■

משפט 4.4 סגירות שפות רגולריות תחת הפרש

אם L_1, L_2 שפות רגולראיות אז $L_1 \Delta L_2 = (L_1 \backslash L_2) \cup (L_2 \backslash L_1)$ היא שפה רגולרית.

הוכחה:

- L_1 שפה רגולרית $\Longleftarrow$ $\exists$ אוטומט $A_1 = (Q_1, \Sigma_1, \delta_1, q_0, F_1)$ שמקבלת אותה.
- L_2 שפה רגולרית $\Longleftarrow$ $\exists$ אוטומט $A_2 = (Q_2, \Sigma_2, \delta_2, p_0, F_2)$ שמקבלת אותה.
- קיים אוטומט דטרמיניסטי A המקבל את השפה $L_1 \cup L_2$, והוא האוטומא המכפלה הבא:

$$A = (Q_1 \times Q_2, \Sigma, \delta, (q_0, p_0), (F_1 \times Q_2 \backslash F_2) \cup (Q_1 \backslash F_1 \times F_2)) \ .$$

כאשר הפונקציית המעברים δ מוגדרת כמו באוטומט מכפלה:

$$\delta((q,p),\sigma) = (\delta_1(q,\sigma), \delta_2(p,\sigma)) \ .$$

הוכחת נכונות.

אפשר להוכיח כי

$$A = (Q_1 \times Q_2, \Sigma, \delta, (q_0, p_0), (F_1 \times Q_2 \backslash F_2) \cup (Q_1 \backslash F_1 \times F_2)) \ .$$

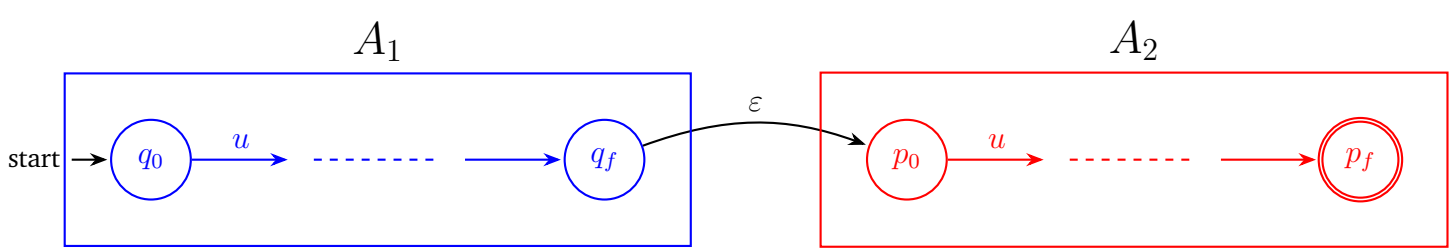
מקבל את השפה $L_1 \Delta L_2 = (L_1 \backslash L_2) \cup (L_2 \backslash L_1)$ באותו אופן של ההוכחה שקיים אוטומט המקבל את השפה ■

משפט 4.5 סגירות שפות רגולריות תחת שרשור

אם L_1, L_2 שפות רגולראיות אז $L_1 L_2$ היא שפה רגולרית.

הוכחה:

- L_1 שפה רגולרית $\Longleftarrow$ $\exists$ אוטומט A_1 שמקבלת אותה.
- L_2 שפה רגולרית $\Longleftarrow$ $\exists$ אוטומט A_2 שמקבלת אותה.
- נבנה אוטומט לא דטרמיניסטי N עם מעברי- ε המקבל את השפה $L_1 L_2$ באופן הבא.



אם $A_1 = (Q_1, \Sigma_1, \delta_1, q_0, F_1)$ ו- $A_2 = (Q_2, \Sigma_2, \delta_2, p_0, F_2)$ אזי

$$N = (Q_1 \cup Q_2, \Sigma_1 \cup \Sigma_2, \Delta, q_0, F_2)$$

כאשר

$$\Delta(q,\sigma) = \begin{cases} \delta_1(q,\sigma) & : q \in Q_1 \backslash F \ , \\ \delta_2(q,\sigma) & : q \in Q_2 \backslash F \ , \\ \{\delta_1(q_f,\sigma), p_0\} & : q \in F \ . \end{cases}$$

הוכחת נכונות.

כיוון 1:

אם $w \in L_1L_2$

$\Longleftarrow$ $w = uv$ כך ש: $u \in L_1, v \in L_2$.

$\Longleftarrow$ A_1 מקבל u ו- A_2 מקבל v.

$\Longleftarrow$ קיים מסלול חישוב של A_1 על u שמסיים במצב q_f וקיים מסלול חישוב של A_2 על v שמסיים במצב p_f.

$\Longleftarrow$ קיים מסלול חישוב של N על w שמסיים במצב p_f.

$\Longleftarrow$ N מקבל w.

כיוון 2:

אם $w \notin L_1L_2$

$\Longleftarrow$ לא קיימות מילים u, v כך ש: $w = uv$ כאשר $u \in L_1, v \in L_2$.

$\Longleftarrow$ לכל חלוקה $w = uv$: או A_1 לא מקבל את u או A_2 מקבל את v.

$\Longleftarrow$ לא קיים מסלול חישוב של A_1 על u שמסיים במצב q_f או לא וקיים מסלול חישוב של A_2 על v שמסיים במצב p_f.

$\Longleftarrow$ N על w שמסיים במצב p_f.

$\Longleftarrow$ N לא מקבל את w.

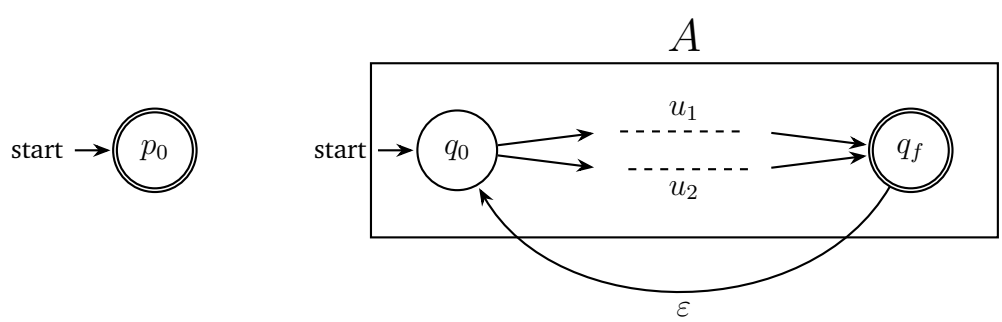
■

משפט 4.6 סגירות שפות רגולריות תחת סגור קליין

אם L שפות רגולראיות אז L^* היא שפה רגולרית.

הוכחה:

- L שפה רגולרית $\Longleftarrow$ $\exists$ אוטומט A שמקבל אותה.
- נבנה אוטומט לא דטרמיניסטי N עם מעברי- ε המקבל את השפה L^* באופן הבא.



אם $A=(Q,\Sigma,\delta,q_0,F)$ אזי

$$N=(Q\cup\{p_0\},\Sigma,\Delta,q_0,F\cup\{p_0\})$$

כאשר

$$\Delta(q,\sigma)=\begin{cases}\delta(q,\sigma) & :q\notin F\ ,\\ \{\delta(q_f,\sigma),q_0\} & :q\in F\ ,\\ p_0 & :q=p_0\ .\end{cases}$$

הוכחת נכונות.

ראשית נציין שאם $w=\varepsilon$ אז כל הרצה של N על w שמסיים במצב המקבל p_0.

כעת נוכיח את הטענה עבור $w\neq\varepsilon$ באופן הבא.

<u>כיוון 1:</u>

אם $w\in L^*$ וגם $w\neq\varepsilon$

$\Longleftarrow$ קיים $n\in\mathbb{N}^+$ כך ש-

$$w=w_1\cdots w_n$$

כאשר $w_i\in L$ לכל $1\leqslant i\leqslant n$,

$\Longleftarrow$ לכל $1\leqslant i\leqslant n$, A מקבל את w_i.

$\Longleftarrow$ לכל $1\leqslant i\leqslant n$, קיים מסלול חישוב של A על w_i שמסיים במצב $q_f\in F$.

$\Longleftarrow$ קיים מסלול חישוב של N על $w=w_1\cdots w_n$ שמסיים במצב $q_f\in F$.

$\Longleftarrow$ לכל $n\in\mathbb{N}^+$, N מקבל את $w=w_1\cdots w_n$.

<u>כיוון 2:</u>

אם $w\notin L^*$ וגם $w\neq\varepsilon$

$\Longleftarrow$ לא קיים $n\in\mathbb{N}^+$ כך ש-

$$w=w_1\cdots w_n$$

כאשר $w_i\in L$ לכל $1\leqslant i\leqslant n$,

$\Longleftarrow$ לכל $n\in\mathbb{N}^+$, לכל חלוקה

$$w=w_1\cdots w_n\ ,$$

קיים $1\leqslant i\leqslant n$ עבורו A לא מקבל את w_i.

$\Longleftarrow$ קיים $1\leqslant i\leqslant n$, עבורו לא קיים מסלול חישוב של A על w_i שמסיים במצב $q_f\in F$.

$\Longleftarrow$ לא סקיים מסלול חישוב של N על $w=w_1\cdots w_n$ שמסיים במצב $q_f\in F$.

$\Longleftarrow$ N לא מקבל את $w=w_1\cdots w_n$.

■

ניתן לשלב מספר סגירויות ביחד כמו בדוגמה הבאה.

דוגמה 4.3

תהיינה L_1, L_2, L_3 שפות רגולראיות. הוכיחו כי השפה

$$\left((L_1 \cup L_2)^*\right) \cdot L_3$$

היא גם שפה רגולרית.

פתרון:

$\Longleftarrow$ $L_1 \cup L_2$ רגולרית ממשפט סגירות של שפות רגולריות תחת איחוד.

$\Longleftarrow$ מכיוון ש- $L_1 \cup L_2$ רגולרית אז $(L_1 \cup L_2)^*$ רגולרית ממשפט סגירות של שפות רגולריות תחת סגור קליין.

$\Longleftarrow$ מכיוון ש- $(L_1 \cup L_2)^*$ וגם L_3 רגולרית (נתון) אז השרשור $(L_1 \cup L_2)^* L_3$ גם שפה רגולרית ממשפט סגירות של שפות רגולריות תחת שרשור.

משפט 4.7

כל שפה סופית היא רגולרית.

הוכחה:

- עבור כל מילה בודדת קיים אוטומט A המקל אותה (שיטת השלד).
- כל מילה בודדת היא שפה רגולרית.
- שפה סופית היא סדרה סופיתשל איחודים של שפות כאלו, וממשפט סגירות שפות רגלריות תחת איחוד, שפה סופית רגולרית.

■

משפט 4.8

אם L שפה רגולרית אז גם L^r שפה רגולרית, כאשר L^r היא השפת ההיפוך L שמוגדרת:

$$L^r = \left\{w^r \middle| w \in L\right\} \ .$$

הוכחה: L רגולרית אז קיים אוטומט A שמקבל L. נבנה אוטומט B שמקבל L^r באופן הבא.
אם

$$A = (Q_A, \Sigma, \delta_A, q_{A0}, F_A)$$

אז

$$B = (Q_B, \Sigma, \delta_B, Q_{B0}, F_B)$$

כאשר

- $Q^B = Q^A$,
- $Q_{B0} = F_A$ ו- $F_B = \{q_{A0}\}$,
- $$\forall p \in Q_B \ : \ \forall \sigma \in \Sigma : \qquad \delta_B\left(p, \sigma\right) = \left\{q \ \middle| \ \delta_A\left(q, \sigma\right) = p\right\} \ .$$

הוכחת נכונות

כיוון 1:

אם $w \in L^R$

$\Leftarrow$ $w^R \in L$

$\Leftarrow$ A מקבל את w^R.

$\Leftarrow$ קיים מסלול חישוב של A על w^R שמתחיל במצב ההתחלתי q_{A0} ומסתיים במצב $q_f \in F_A$.

$\Leftarrow$ על פי הגדרת המעברים ב-B, קיים מסלול חישוב הפוך של B על w שמתחיל במצב $q_f \in Q_{B0}$ ומסתיים במצב $q_{A0} \in F_B$.

$\Leftarrow$ הרצה של B על w מסתיימת במצב מקבל.

$\Leftarrow$ B מקבל את w.

כיוון 2:

אם $w \notin L^R$

$\Leftarrow$ $w^R \notin L$

$\Leftarrow$ A לא מקבל את w^R.

$\Leftarrow$ לא קיים מסלול חישוב של A על w^R שמתחיל במצב q_{A0} ומסתיים במצב $q_f \in F_A$.

$\Leftarrow$ על פי הגדרת המעברים ב-B, לא קיים מסלול חישוב של B על w שמתחיל באחד ממצבי ההתחלה ב-Q_{B0} ומסתיים במצב המקבל $q_{A0} \in F_B$.

$\Leftarrow$ הרצה של B על w מסתיימת במצב שאינו מצב מקבל.

$\Leftarrow$ B לא מקבל את w.

■

4.3 סיכום

משפט 4.9 סגירות שפות רגולריות

תהיינה L_1, L_2, L שפות רגולריות. אזי השפות הבאות גם רגולריות:

1) החיתוך: $L_1 \cap L_2$

2) האיחוד: $L_1 \cup L_2$

3) ההפרש: $L_1 \backslash L_2$

4) ההפרש סימטרי: $L_1 \Delta L_2$

5) השרשור: $L_1 L_2$

6) הסגור קליין:

7) השרשור: L^*

8) ההיפוך: L^r

9) הרישא: $\text{prefix}(L) = \{x \in \Sigma^* \mid \exists y \in \Sigma^* : \ xy \in L\}$

המצבים המקבלים לכל פעולה (חיתוך, איחוד, הפרש והפרש סימטרי) מתוארים בטבלה למטה:

פעולה	שפה	מצבים מקבלים
חיתוך	$L_1 \cap L_2$	$F_1 \times F_2$
איחוד	$L_1 \cup L_2$	$(F_1 \times Q_2) \cup (Q_1 \times F_2)$
הפרש	$L_1 \backslash L_2$	$F_1 \times (Q_2 \backslash F_2)$
הפרש	$L_1 \Delta L_2$	$(F_1 \times (Q_2 \backslash F_2)) \cup ((Q_1 \backslash F_1) \times F_2)$

4.4 ביטויים רגולריים

הגדרה 4.3 ביטוי רגולרי

מחרוזת r הינה ביטוי רגולרי אם היא מאחת הצורות הבאות:

בסיס האינדוקציה:

- σ, כאשר σ אות בודדת בא"ב.
- ε
- $\varnothing$

פעולות:

- $r_1 r_2$, כאשר r_1 ו-r_2 ביטויים רגולריים.
- $r_1 | r_2$, כאשר r_1 ו-r_2 ביטויים רגולריים.
- r_1^*, כאשר r_1 ביטוי רגולרי.
- (r_1), כאשר r_1 ביטוי רגולרי.

דוגמה 4.4

א) הביטוי הרגולרי

$$(0|1)^*000$$

היא שפת כל המחרוזות מעל $\Sigma = \{0,1\}$ המסתיימות ב- 000.

ב) הביטוי הרגולרי

$$(0|1)^*000(0|1)^*$$

היא שפת כל המילים מעל $\Sigma = \{0,1\}$ המכילות בתוכן את הצירוף 000 (בכל מקום במחרוזת).

ג) הביטוי הרגולרי

$$\left((a|b)^*aaa \;\middle|\; bb\,(a|b)^*\right)$$

היא שפת כל המילים מעל $\Sigma = \{a, b\}$ המסתיימות ב- aaa או מתחילות ב- bb.

ד) הביטוי הרגולרי $(0|1)^*(01)^{99}$ מתאר את שפת כל המילים מעל האלפבית $\Sigma = \{0, 1\}$ המסתיימות במחרוזת $(01)^{99}$.

ה) הביטוי הרגולרי $0^*1^*2^*$ מתאר את שפת כל המילים מעל האלפבית $\Sigma = \{0, 1, 2\}$ המתחילות במספר כלשהו של אפסים ולאחר מכן מספר כלשהו של 1-ים ולבסוף מספר כלשהו של 2 -ים.

ו) הביטוי הרגולרי $00^*11^*22^*$ מתאר את שפת כל המילים מעל האלפבית $\Sigma = \{0, 1, 2\}$ המתחילות בסדרת אפסים ולאחר מכן סדרת 1-ים ולבסוף סדרת 2 -ים. כל אחת מהסדרות מכילה לפחות איבר אחד.

דוגמה 4.5

הוכיחו כי $(0|01)^*$ מתאר את שפת כל המילים מעל הא"ב $\{0, 1\}$ אשר מתחילות ב- 0 אבל אינן מכילות 11.

פתרון:

הגדרה 4.4

יהי r ביטוי רגולרי. **השפה של** r , נסמן $L(r)$ הינה השפה הבאה:

בסיס האינדוקציה:

- אם $r = \sigma$, כאשר σ אות בודדת בא"ב אזי $L(r) = \{\sigma\}$
- אם $r = \varepsilon$ אזי $L(r) = \{\varepsilon\}$.
- אם $r = \varnothing$ אזי $L(r) = \varnothing$.

פעולות:

- אם $r = r_1 \circ r_2$ אזי $L(r) = L(r_1) \circ L(r_2)$
- אם $r = r_1|r_2$ אז $L(r) = L(r_1) \cup L(r_2)$.
- אם $r = r_1^*$ אז $L(r) = L(r_1)^*$.
- אם $r = (r_1)$ אז $L(r) = L(r_1)$.

דוגמה 4.6 מילים הכוללות bb

מצאו ביטוי רגולרי לשפת כל המילים מעל $\{a, b\}$ שכוללות את הרצף bb.

פתרון:

- המילה הבסיסית ביותר היא bb.

- לפני המילה הבסיסית יכול להופיע רצף כלשהו של אותיות a, b.
- אחרי המילה הבסיסית יכול להופיע רצף כלשהו של אותיות a, b.
- לכן הביטוי הרגולרי הוא:
$$r = (a|b)^* bb(a|b)^* \ .$$

דוגמה 4.7 b במקומות הזוגיים

מצאו ביטוי רגולרי לשפת כל המילים מעל $\{a, b\}$ שבהן במיקומים הזוגיים מופיע b.

פתרון:

- האות הראשונה יכולה להיות או a או b כי המקום הראשון הוא מקום איזוגי.
- האות השניה חייבת להיות b כי מקום 2 הוא מקום זוגי.
- וניתן לחזור על זה שוב ושוב.
- לכן הביטוי הרגולרי הוא:
$$r = ((a|b)b)^* \ .$$
- אם רוצים לאפשר גם מילים באורך איזוגי, נוסיף בסוף אפשרות לכל אחת מהאותיות, בכדי לאפשר אורך איזוגי, וגם נוסיף בסוף ε, למקרה שרוצים אורך זוגי.
- לכן הביטוי הרגולרי הוא:
$$r = ((a|b)b)^* \, (a|b|\varepsilon) \ .$$

דוגמה 4.8

מצאו ביטוי רגולרי לשפת כל המילים מעל $\{a, b\}$ שאורכן מתחלק ב- 3 או 2 ללא שארית.

פתרון:

$$\Big(\ ((a|b)(a|b))^* \ \Big| \ ((a|b)(a|b)(a|b))^* \ \Big)$$

דוגמה 4.9

מצאו ביטוי רגולרי לשפת כל המילים מעל $\{a, b\}$ שלא מסתיימות ברצף ba.

פתרון:

$$(a|b|\varepsilon) \ \Big| \ \Big(\ (a|b)^*(aa|b) \ \Big)$$

דוגמה 4.10

מצאו ביטוי רגולרי לשפת כל המילים מעל $\{a, b\}$ שמכילות את הרצף aa וגם את הרצף bb.

פתרון:

$$(\ (a|b)^*aa(a|b)^*bb(a|b)^*\ )\ \Big|\ (\ (a|b)^*bb(a|b)^*aa(a|b)^*\ )$$

הגדרה 4.5 דרגה של ביטוי רגולרי

לכל ביטוי רגולרי נתאים דרגה על פי המתווה הבא:

א) ביטויים רגולריים בעלי דרגת מורכבות 0: $\varnothing, \varepsilon, a$.

ב) אם r, s ביטויים רגולריים בעלי דרגת מורכבות n, m בהתאמה, אז דרגת המורכבות של rs ו- $r|s$ תהיה $\max\{m, n\} + 1$.

ג) אם r ביטוי רגולרי בעלי דרגת מורכבות n, אז דרגת המורכבות של r^* תהיה $n + 1$.

סימון: את דרגת המורכבות של ביטוי רגולרי r נסמן ע"י $\deg[r]$.

$$\begin{aligned}
\deg[r^*] &= \deg[r] + 1\ , \\
\deg[rs] &= \max\left(\deg[r], \deg[s]\right) + 1\ , \\
\deg[r|s] &= \max\left(\deg[r], \deg[s]\right) + 1\ .
\end{aligned}$$

דוגמה 4.11

א)

$$\begin{aligned}
\deg\left[(a|ab)^*\right] &= \deg\left[(a|ab)\right] + 1 \\
&= \max\left(\deg\left[a\right], \deg\left[ab\right]\right) + 2 \\
&= \max\left(0, \max\left(\deg\left[a\right], \deg\left[b\right]\right) + 1\right) + 2 \\
&= \max\left(0, \max\left(0, 0\right) + 1\right) + 2 \\
&= 3\ .
\end{aligned}$$

ב)

$$\begin{aligned}
\deg\left[bba\right] &= \max\left(\deg\left[b\right], \deg\left[ba\right]\right) + 1 \\
&= \max\left(0, \max\left(\deg\left[b\right], \deg\left[a\right]\right) + 1\right) + 1 \\
&= \max\left(0, \max\left(0, 0\right) + 1\right) + 1 \\
&= 2\ .
\end{aligned}$$

ג)

$$\begin{aligned}
\deg\left[(bbaa)^*\right] &= \deg\left[bbaa\right] + 1 \\
&= \max\left(\deg\left[b\right], \deg\left[baa\right]\right) + 2 \\
&= \max\left(0, \max\left(\deg\left[b\right], \deg\left[aa\right]\right) + 1\right) + 2 \\
&= \max\left(0, \max\left(0, \max\left(\deg\left[a\right], \deg\left[a\right]\right) + 1\right) + 1\right) + 2 \\
&= \max\left(0, \max\left(0, \max\left(0, 0\right) + 1\right) + 1\right) + 2 \\
&= 4
\end{aligned}$$

ד)

$$\begin{aligned}
\deg\left[(a|ab)^*(bbaa)^*\right] &= \max\left(\deg\left[(a|ab)^*\right], \deg\left[(bbaa)^*\right]\right) + 1 \\
&= \max\left(3, 4\right) + 1 \\
&= 5\ .
\end{aligned}$$

הגדרה 4.6 אוטומט סופי מוכלל $GNFA$

הגדרת החמשת מרכיבים

אוטומט מוכלל הוא אוטומט לא דטרמיניסטי עם מעברי- ε:

$$G = (Q, \Sigma, \delta, q_s, q_f)$$

כשאר:

1) Q: קבוצת **מצבים** סופית.

2) Σ: **האלפבית** של הקלט.

3) δ: **הפונקצית המעברים**,

$$\delta : (Q \backslash \{q_f\}) \times (Q \backslash \{q_s\}) \to R$$

כאשר R האוסף של כל הביטויים הרגולריים.

δ מקבל כקלט שני מצבים q_i, q_j ומחזירה הביטויי הרגולרי שמסמן את הצלע היוצאת מ- q_i ונכנסת ל- q_j.

4) q_s המצב ההתחלתי היחיד.

5) q_f המצב המקבל היחיד.

הכללי הביצוע של $GNFA$

$GNFA$ מתבצע לפי הכללים הבאים:

- כל צלע מסומנת על ידי ביטוי רגולרי באלפבית Σ.
- קיים מצב התחלתי אחד ויחיד q_s אשר אליו לא נכנסת שום צלע.
  קיימת צלע היוצאת מ- q_s לכל מצב אחר.
- קיים מצב קבלה אחד ויחיד q_f אשר ממנו לא יוצאת שום צלע. על כל מצב אחר קיימת צלע ממנו הנכנסת למצב המקבל.
- קיימת קשת בין כל שני קדקודים, וקיימת וקשת בין כל קודקוד לעצמו, למעט העובדה שאין כניסות למצב ההתחלתי ואין יציאות מהמצב המקבל.
- המעבר ממצב q_i למצב q_j הוא אפשרי רק על ידי קליטה של מחרוזת התואמת את הביטוי הרגולרי המסמן את הצלע היוצאת מ- q_i ל- q_j.
- אם קיימת צלע היוצאת ממצב q_i ונכנסת למצב q_j המסומנת ע"י ביטוי רגולרי r אז

$$\delta(q_i, q_j) = r \ .$$

קבלה של מילה ע"י $GNFA$

קבלה של מילה w ע"י אוטומט $GNFA$ מוגדרת באופן הבא. יהי $GNFA\ G = (Q, \Sigma, \delta, q_s, q_f)$.
אומרים כי G מקבל מילה $w \in \Sigma^*$ אם ורק אם קיים פירוק $w = w_1 w_2 \cdots w_k$, כאשר לכל $1 \leqslant i \leqslant k$: $w_i \in \Sigma^*$, וקיימים סדרת מצבים $q_0, q_1, \ldots, q_k \in Q$ כך ש:

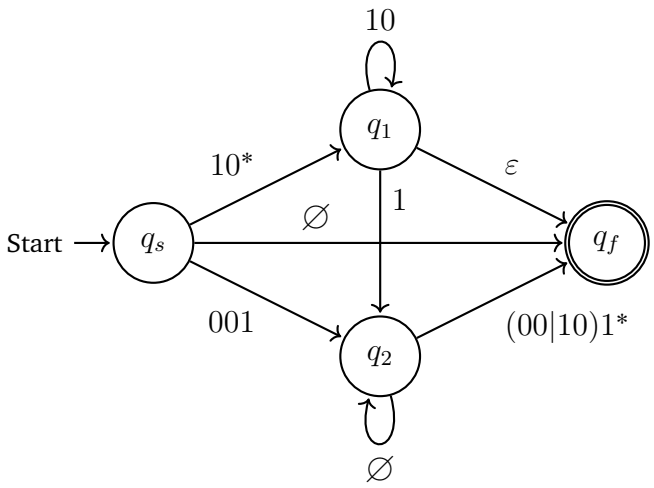
1) $q_0 = q_s$ הוא המצב ההתחלתי,

2) $q_k = q_f$ הוא המצב המקבל,

3) לכל $1 \leqslant i \leqslant k$: $w_i \in L(r_i)$ כאשר $r_i = \delta\left(q_{i-1,q_i}\right)$.
כלומר r_i הוא הביטוי על הצלע מ- q_{i-1} ל- q_i.

דוגמה 4.12

יהי M האוטומט אשר דיאגרמת המעבר שלו מוצגת בשרטוט למטה.



דוגמה לכמה מחרוזות המתקבלות ע"י האוטומט M הן:

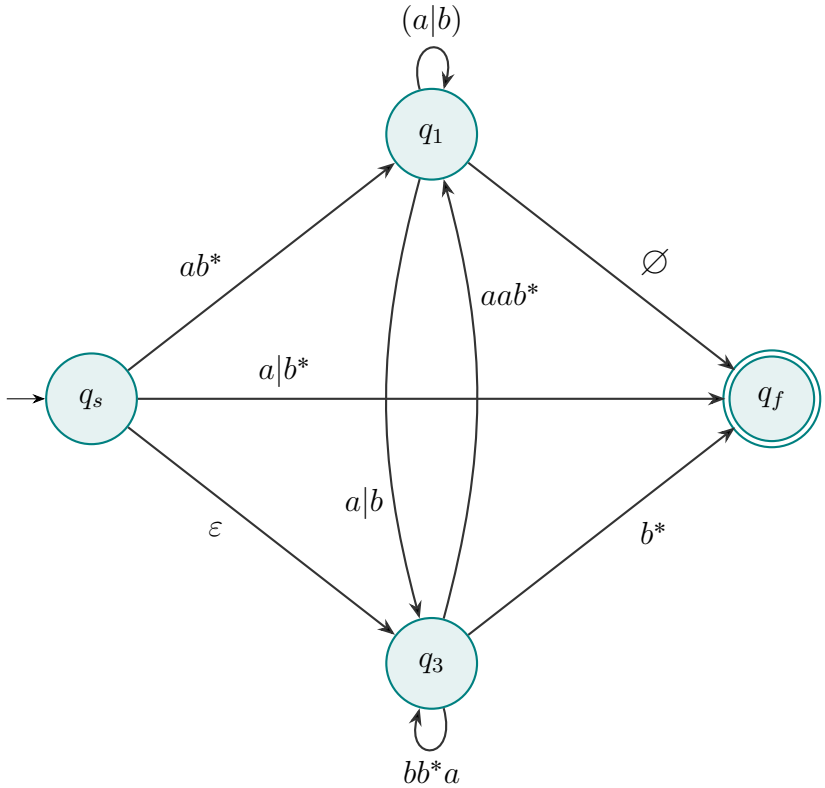
$$00100, \quad 100000, \quad 10001010, \quad 1110111 \ .$$

למשל סדרת חישוב עבור $w = 1001010$ היא:

$$q_s \xrightarrow{1000} q_1 \xrightarrow{10} q_1 \xrightarrow{10} q_1 \xrightarrow{\varepsilon} q_f \ .$$

דוגמה 4.13

יהי M האוטומט אשר דיאגרמת המעבר שלו מוצגת בשרטוט למטה.



דוגמה לכמה מחרוזות המתקבלות ע"י האוטומט M הן:

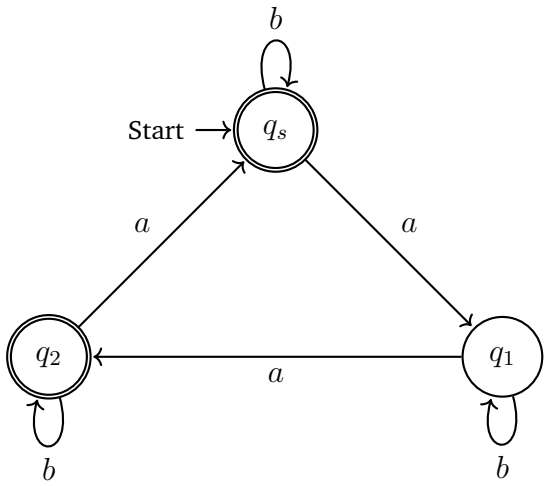
$$a, \quad bbbbbb, \quad bbbbbbbbabbb, \quad aba \ .$$

למשל סדרת חישוב עבור $w = abbabba$ היא:

$$q_s \xrightarrow{ab} q_1 \xrightarrow{b} q_1 \xrightarrow{a} q_3 \xrightarrow{bba} q_3 \xrightarrow{bb} q_f \ .$$

דוגמה 4.14 מעבר מאוטומט סופי לאוטומט סופי מוכלל

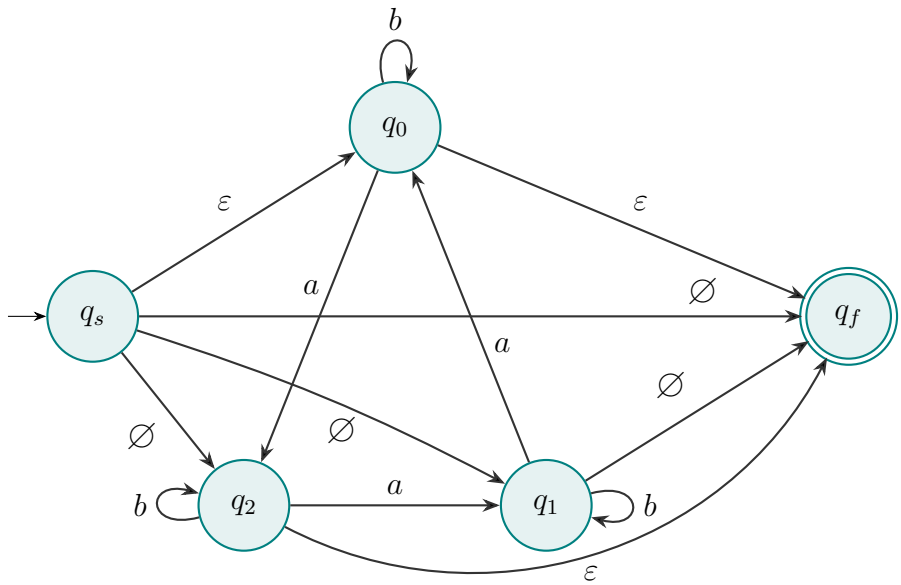
נתון האוטומט הסופי הבא.



בנו אוטומט סופי מוכלל שקול.

פתרון:
נעשה זאת בשלבים:

- ניצור מצב התחלתי יחיד.
- ניצור מצב מקבל יחיד.
- נוסיף קשתות בין כל זוג מצבים, שביניהם אין צלע, צלע מסומנת ע"י הקבוצה הריקה.



משפט 4.10

לכל אוטומט סופי $A \in$ קיים אוטומט סופי מוכלל B שקול. כלומר:

$$L(A) = L(B) \ .$$

הוכחה: ■

הגדרה 4.7 שיטת דילול מצבים

נתון אוטומט סופי מוכלל G. האלגוריתם הבא מקבל כקלט G ונותן כפלט אוטומט סופי מוכלל חדש G', עם רק שני מצבים.

$CONVERT(G) =$ " על קלט $G = (Q, \Sigma, \delta, q_s, q_f)$, מבצע השיטת הרקורסיבית הבאה:

שלב 1) יהי k מספר המצבים של G.

שלב 2) אם $k = 2$ אזי G מורכב ממצב התחלתי q_s, מצב המקבל q_f, וצלע יחיד מ- q_s ל- q_f המסומנת עם ביטוי רגולרי, r. $\Leftarrow$ מחזיר r.

שלב 3) אם $k > 2$, בוחר מצב $q' \in Q$ כאשר $q' \neq q_s$ וגם $q' \neq q_f$.
יהי G' ה- $GNFA$ הבא: $G' = (Q', \Sigma, \delta', q_s, q_f)$ כאשר

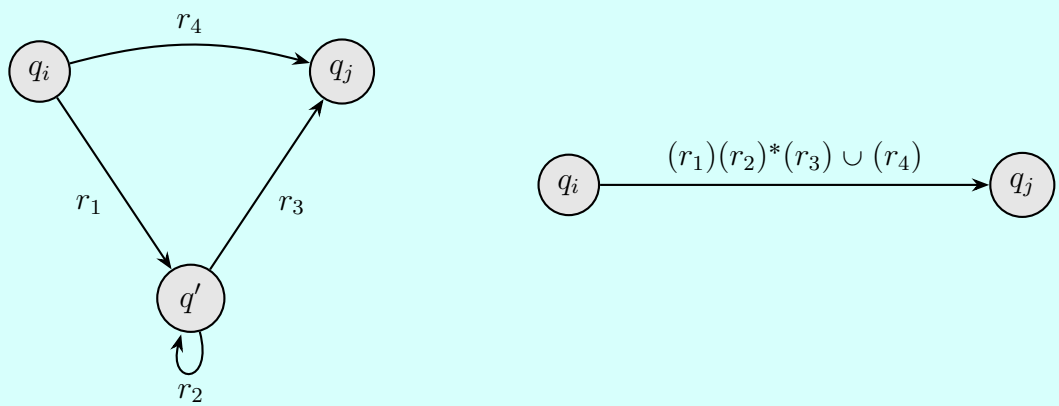
$$Q' = Q \backslash \{q'\}$$

ולכל $q_i \in Q' \backslash \{q_f\}$ ולכל $q_j \in Q' \backslash \{q_s\}$,

$$\delta'(q_i, q_j) = (r_1)(r_2)^*(r_3) \cup (r_4) \ ,$$

כאשר

$$r_1 = \delta(q_i, q'), \quad r_2 = \delta(q', q'), \quad r_3 = \delta(q', q_j), \quad r_4 = \delta(q', q') \ .$$

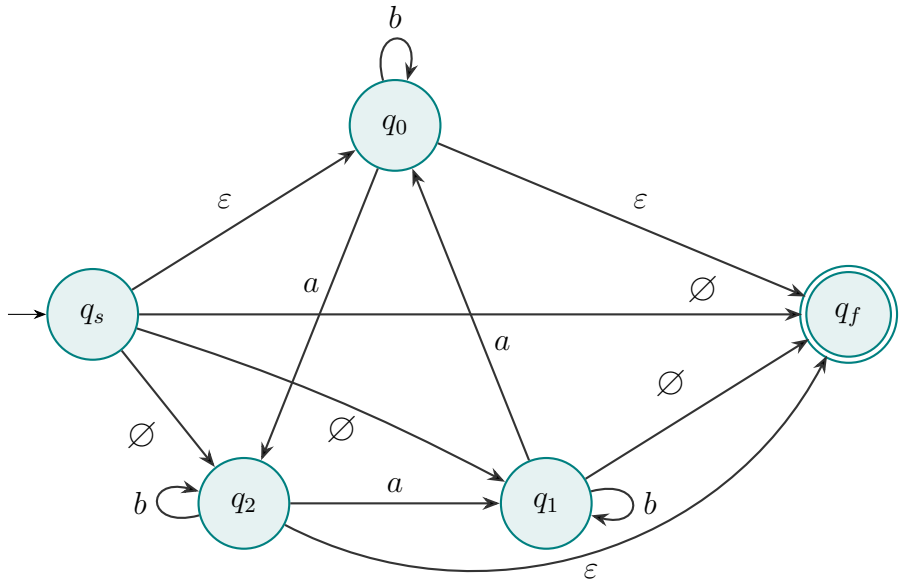


לפני אחרי

שלב 4) $CONVERT(G')$. "

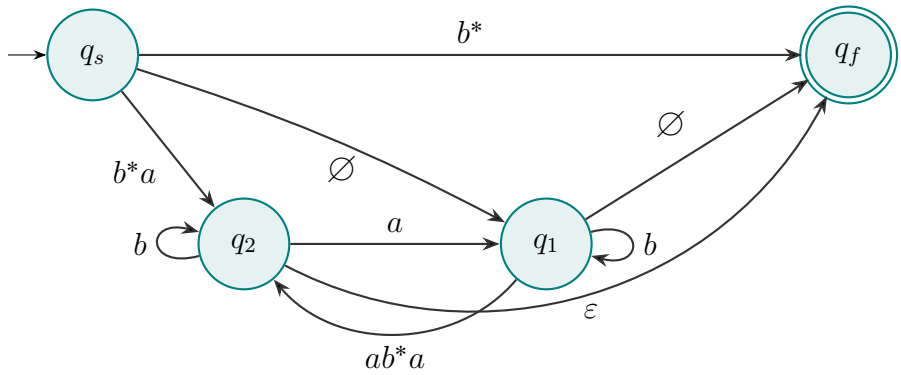
דוגמה 4.15

נתון ה- $GNFA$ הבא:

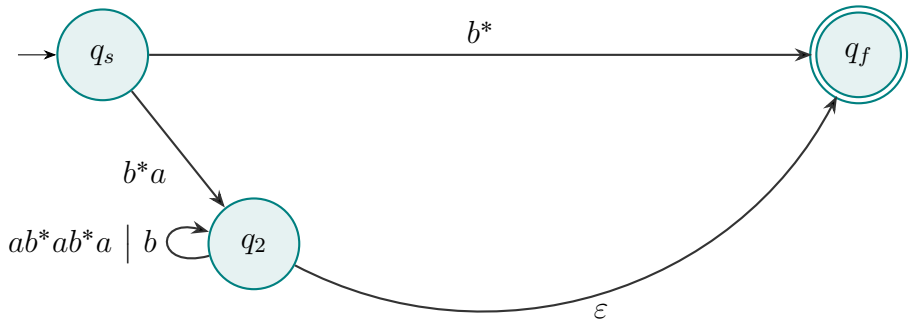


נבצע האלגוריתם דילול מצבים $CONVERT(G)$ עליו כדי לקבל אוטומט G' עם שני מצבים שקול.

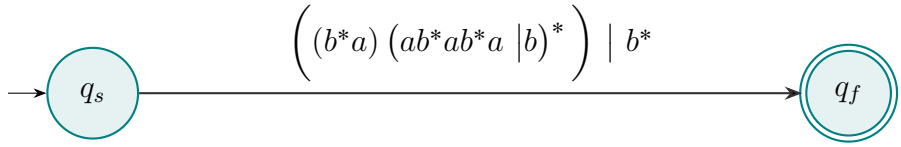
<u>שלב 1</u>



שלב 2

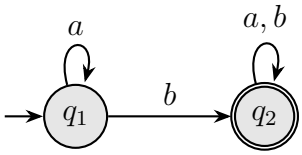


שלב 3



דוגמה 4.16

נתון האוטומט הבא:

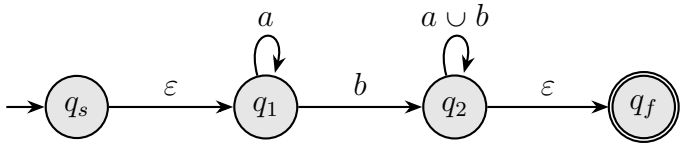


א) בנו אטוטמט מוכלל שקול.

ב) ע"י שיטת הדילול מצבים בנו אוטומט עם שני מצבים שקול.

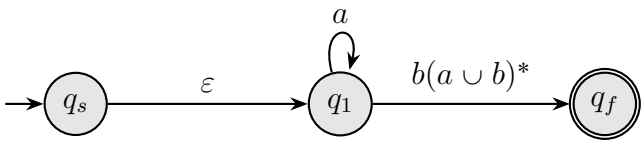
פתרון:

סעיף א)

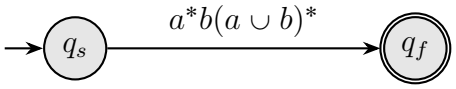


סעיף ב)

שלב 1

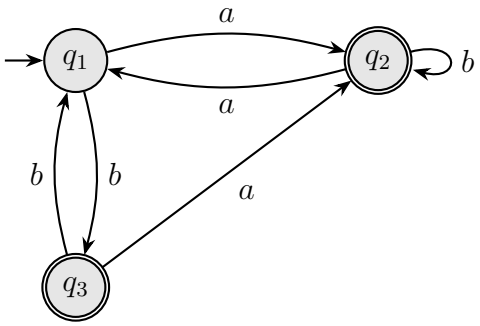


שלב 2



דוגמה 4.17

נתון האוטומט הבא:

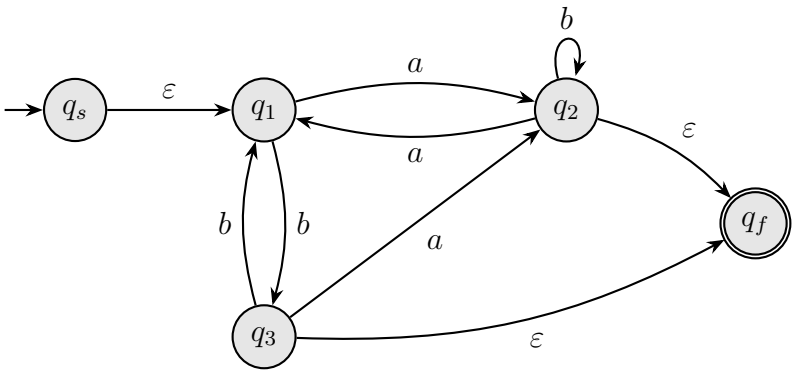


א) בנו אטוטמט מוכלל שקול.

ב) ע"י שיטת הדילול מצבים בנו אוטומט עם שני מצבים שקול.

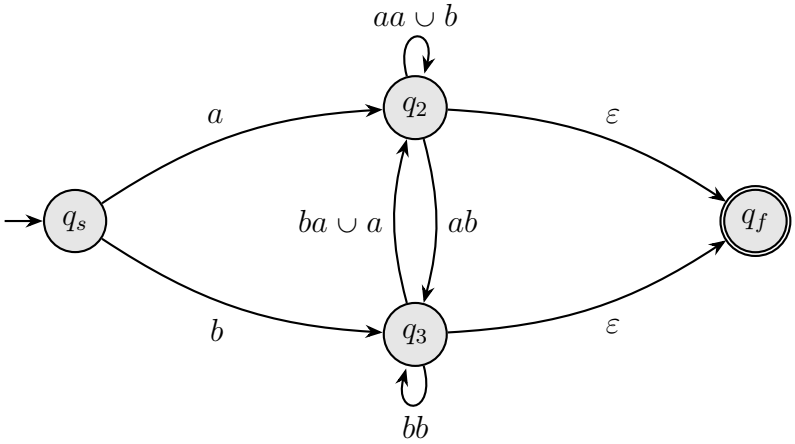
פתרון:

סעיף א)

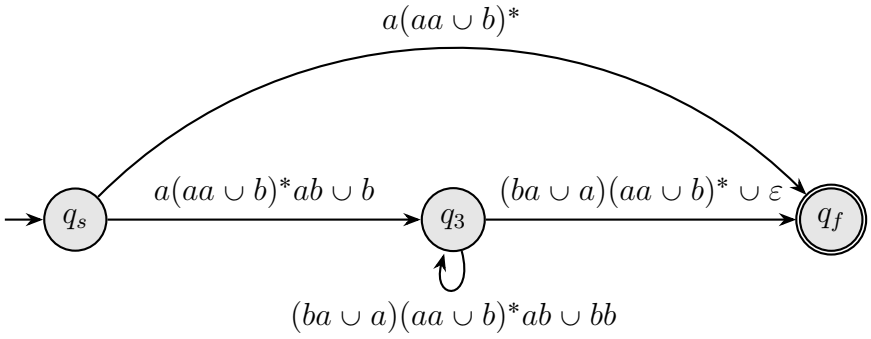


סעיף ב)

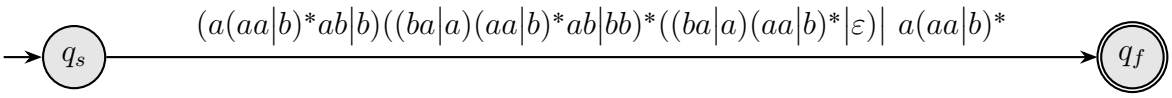
שלב 1



שלב 2

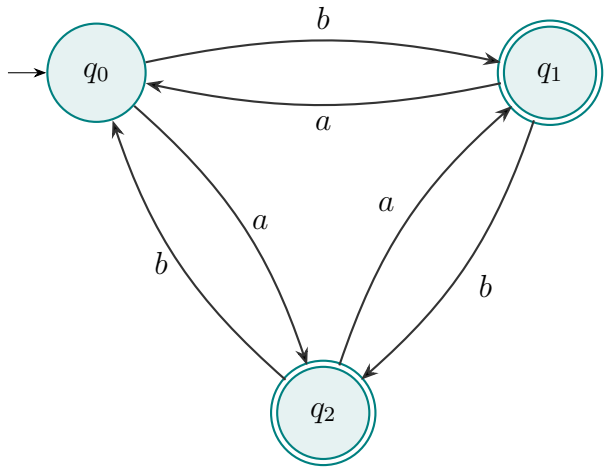


שלב 3



דוגמה 4.18

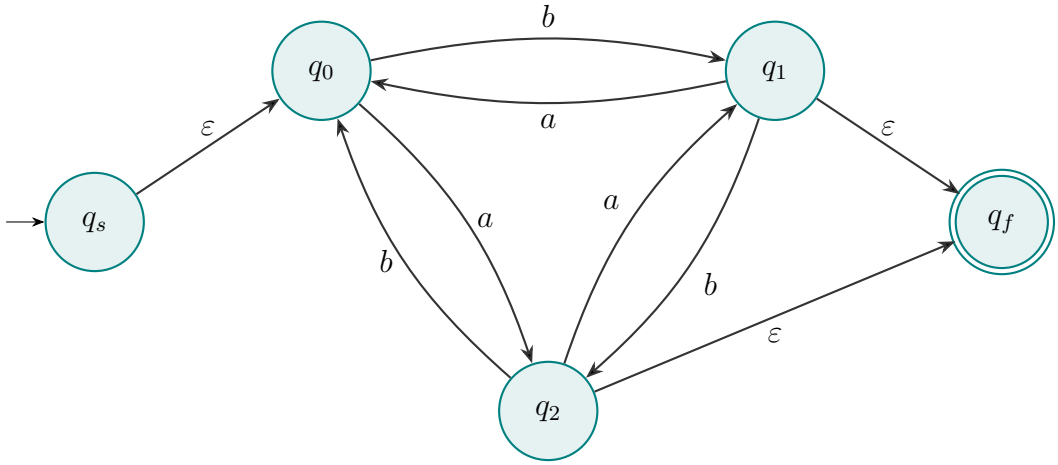
יהי A האוטומט הבא:



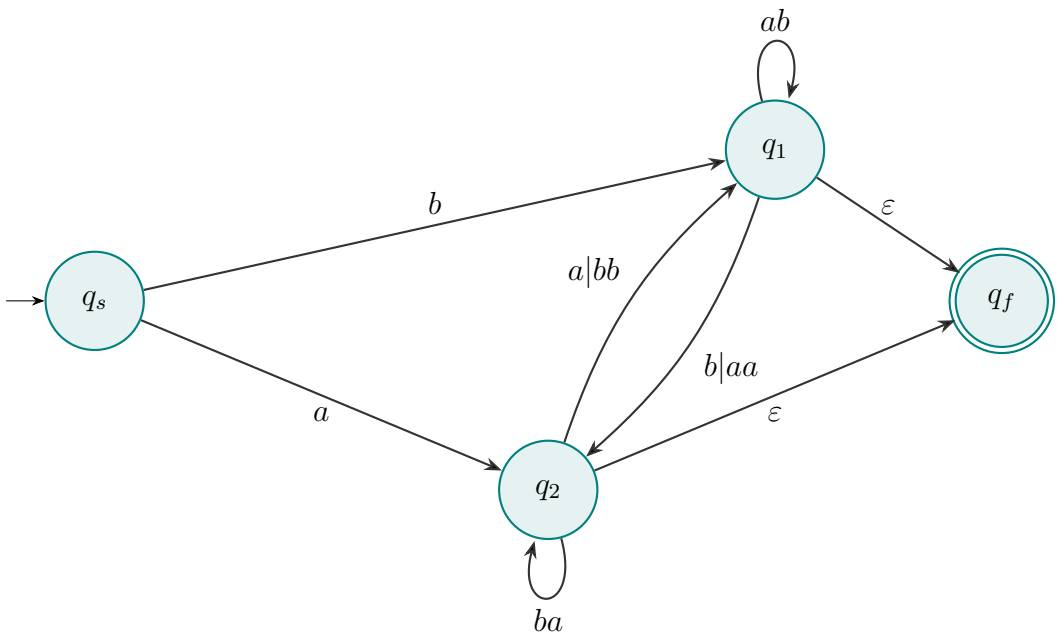
מצאו ביטוי רגולרי r כך ש- $L(r) = L(A)$.

פתרון:

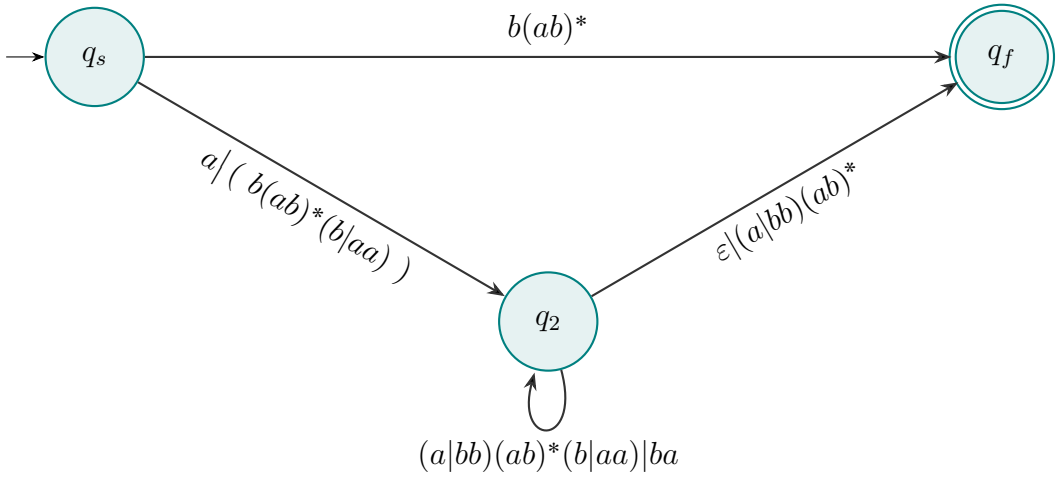
ראשית ניצור אוטומט סופי מוכלל שקול:



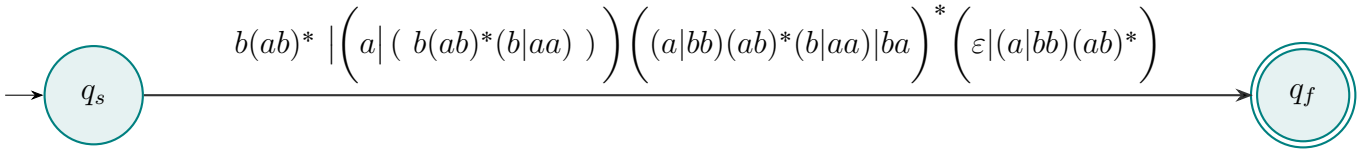
דילול q_0



דילול q_1



דילול q_2



משפט 4.11

תהי L שפה. L רגולרית אם ורק אם קיים ביטוי רגולרי r שך ש:

$$L = L(r)\ .$$

הוכחה:

<u>כיוון ראשון</u>

נניח קיים ביטוי רגולרי r כך ש- $L = L(r)$. נוכיח כי L רגולרית.

כיוון שני

נניח כי L רגולרית. נוכיח כי קיים ביטוי רגולרי r כך ש- $L = L(r)$. ■

4.5 למת הניפוח

משפט 4.12 למת הניפוח

אם L שפה רגולרית אז קיים מספר טבעי p **(קבוע למת הניפוח)** כך שאם $w \in L$ וגם $|w| \geqslant p$ אז קיים פירוק $w = xyz$ כך שהתנאים הבאים מתקיימים:

1. $|y| > 0$.
2. $|xy| \leqslant p$.
3. לכל $i \geqslant 0$ מתקיים $xy^iz \in L$.

הוכחה:

- יהי $M = (Q, \Sigma, \delta, q_1, F)$ אוטומט סופי דטרמיניסטי שמקבל L ויהי p המספר המצבים של M: כלומר $|Q| = p$.
- תהי $w = \sigma_1 \cdots \sigma_n$ מילה בשפה L כאשר $n \geqslant p$.
- נסמן ב $\{q_1, \ldots, q_{n+1}\}$ את סדרת המצבים ש- M עוברת דרכם במהלך עיבוד המילה w.
- כלומר, $q_{i+1} = \delta(q_i, \sigma_i)$ עבור $1 \leqslant i \leqslant n$.
- אורך סדרה זו הוא $n + 1 \geqslant p + 1$.
- מכיוון שאורך הסדרה הוא $n + 1 \geqslant p + 1$, אך ישנם רק p מצבים שונים ב- M, עקרון שובך היונים (Pigeonhole Principle) מבטיח כי בסדרה יש לפחות מצב אחד שמופיע פעמיים בתוך $p + 1$ המקומות הראשונים בסדרה. נקרא להופעה הראשונה של מצב זה q_j ולהופעה השנייה שלו q_l. מכיוון ש- q_l נמצא בין $p + 1$ המקומות הראשונים, מתקיים $l \leqslant p + 1$.
- נגדיר:
$$x = \sigma_1 \cdots \sigma_{j-1} \ , \qquad y = \sigma_j \cdots \sigma_{l-1} \ , \qquad z = \sigma_l \cdots \sigma_n \ .$$
- מכיוון שהמסלול חישוב של M על w הוא:
$$q_1 \xrightarrow{x} q_j \xrightarrow{y} (q_l = q_j) \xrightarrow{z} q_{n+1} \in F \ ,$$
אזי לכל $i > 0$:
$$q_1 \xrightarrow{x} q_j \xrightarrow{y^i} (q_j) \xrightarrow{z} q_{n+1} \in F \ .$$
- לכן לכל $i > 0$ המילה $xy^iz \in L(M)$.
- מכיוון ש- $j \neq l$ אז $|y| > 0$.

- מכיוון ש- $l \leqslant p+1$ אז $|xy| < p$.

■

דוגמה 4.19

תהי L השפה:

$$L = \left\{ w \in \{a,b\}^* \;\middle|\; \#a_w = \#b_w \right\} \ .$$

הוכיחו כי השפה L לא רגולרית.

פתרון:

- נניח בשלילה ש- L רגולרית.

$\Leftarrow$ קיים p כך שאם $w \in L$ וגם $|w| \geqslant p$ אז $\exists$ פירוק $w = xyz$ כך ש-

1) $|y| > 0$
2) $|xy| \leqslant p$
3) $\forall\, i: \quad xy^iz \in L$.

- תהי $w = a^pb^p \in L$. ז"א $|w| = 2p$.
- ה- p אותיות הראשונות כולן a.
- עבור כל פירוק $w = xyz$, מכיוון $|xy| \leqslant p$, אז המחרוזות x ו- y כוללות אותיות a בלבד.
- אזי כל פירוק $w = xyz$ חייב להיות מהצורה:

$$x = a^j \ , \quad y = a^k \ , \quad z = a^{p-j-k}b^p \ , \qquad (k > 0) \ .$$

- מכיוון ש- $|y| > 0$ אז $k \geqslant 1$.
- למת הניפוח $\Leftarrow$ לכל $i \geqslant 0$: $xy^iz \in L$.

$\Leftarrow$ $w' = xy^2z \in L$.

- מכיוון ש- $k > 0$ אז

$$w' = xy^2z = a^ja^{2k}a^{p-j-k}b^p = a^{p+k}b^p \quad \Rightarrow \quad p+k = \#a_{w'} \neq \#b_{w'} = p \quad \Rightarrow \quad w' \notin L$$

$\Leftarrow$ סתירה לכך שמלמת הניפוח $xy^2z \in L$.

דוגמה 4.20

תהי

$$L = \left\{ a^{2^m} \;\middle|\; m \in \mathbb{N} \right\} \ .$$

כלומר L שפת כל המילים מעל האלפבית $\Sigma = \{a\}$ אשר אורכן הוא 2^m. כלומר:

$$L = \{a, a^2, a^4, a^8, \ldots\} = \{a, aa, aaaa, aaaaaaaa, \ldots\} \ .$$

הוכיחו: L לא רגולרית.

פתרון:

- נניח בשלילה כי L רגולרית.

$\Leftarrow$ קיים p (קבוע למת הניפוח) כך שאם $w \in L$ וגם $|w| \geqslant p$ אז $\exists$ פירוק $w = xyz$ כך ש-

1) $|y| > 0$
2) $|xy| \leqslant p$
3) $\forall\, i: \quad xy^iz \in L$.

- תהי $w = a^{2^p} \in L$.
- לכל פירוק $w = xyz$ המחרוזות x, y, z כוללות אותיות a בלבד.
- על פי התנאי ראשון של למת הניפוח, $|y| > 0$. ז"א אם $|y| = k$ אז $1 \leqslant |y| \leqslant p < 2^p$.
- מהתנאי השלישי של מלמת הניפוח, עבור האורך הניפוח $i = 2$: המילה $xy^2z \in L$.
- מצד שני,

$$|xy^2z| = |xyyz| = |xyz| + |y| = 2^p + k < 2^p + 2^p = 2^{p+1} \ ,$$

ז"א

$$2^p < |xy^2z| < 2^{p+1}$$

ולכן לא קיים טבעי m כך ש: $w' = a^{2^m}$ ולכן $w' = xy^2z \notin L$.

$\Leftarrow$ סתירה לכך שעל פי למת הניפוח: $xy^2z \in L$.

דוגמה 4.21

תהי

$$L = \left\{ w \in \Sigma^* \;\middle|\; w = w^R \right\} \ .$$

השפת הפלינדרומים. הוכיחו: L לא רגולרית.

פתרון:

- נניח בשלילה כי L רגולרית.
- תהי $w = a^pba^p \in L$, כאשר p קבוע למת הניפוח וגם $|w| \geqslant p$.

$\Leftarrow$ כל פירוק $w = xyz$ המקיים התנאי למת הניפוח מהצורה $|y| > 0$ וגם $|xy| \leqslant p$. לכן:

$$x = a^j \ , \quad y = a^k \ , \quad z = a^{p-j-k}ba^p \ , \qquad 0 < k \leqslant p \ .$$

- על פי למת הניפוח, עבור אורך הניפוח $i = 2$ המילה $w' = xy^2z \in L$.
-

$$w' = a^ja^{2k}a^{p-j-k}ba^p = a^{p+k}ba^p \quad \Rightarrow \quad {w'}^R = a^pba^{p+k} \neq w' \quad \Rightarrow \quad w' \notin L \ .$$

$\Leftarrow$ בסתירה לכך שמלמת הניפוח $w' \in L$.

דוגמה 4.22

הוכיחו כי השפה $L = \left\{ a^ib^jc^k \;\middle|\; i, j, k \geqslant 0, j = \max(i, k) \right\}$ אינה רגולרית.

פתרון:

- נניח בשלילה כי L רגולרית.
- אזי קיים n של למת הניפוח כך שאם $w \in L$ וגם $|w| \geqslant n$ אז $\exists$ פירוק $w = xyz$ כך ש-
  1. $|y| > 0$
  2. $|xy| \leqslant n$
  3. $\forall\, i: \quad xy^iz \in L$.
- נבחר $w = a^nb^nc^n$. אזי $|w| > n$ וגם $w \in L$.
- נשים לב שבכל חלוקה כזו חלקי ה- x וה- y מכילים אך ורק אותיות a (כיוון שהם בתוך n האותיות הראשונות במילה) ו- y מכיל לפחות אות אחת.
- נסמן $|y| = k$.
- נסתכל בניפוח $i = 2$.
- מתקיים:
$$w' = xy^2z = xyyz = a^na^kb^nc^n = a^{n+k}b^nc^n \ .$$
- עבור ניפוח זה מתקיים שמספר אותיות b קטן ממסר אותיות a ולכן, $xy^2z \notin L$, בסתירה לכך שלפי למת הניפוח $xy^2z \in L$.
- לכן L אינה רגולרית.