28/10

1

אוטומטים ושפות פורמליות

נושא 1 - מילים ושפות

א) מושגים בסיסיים

א"ב - קבוצת סימנים שיכולה להשתנות לפי הגדרתנו. הקבוצה היא סופית וגם לא ריקה. נסמנה ב-Σ. לדוגמא $\Sigma=\{0,1\}$.

מילה - סדרה סופית של סימנים מתוך א"ב נתון, נסמנה ב-w. בהמשך לדוגמא הקודמת $w=0110$. קוראים את המילה משמאל לימין. (אפסילון) מילה יכולה להיות ריקה, כלומר ללא סימנים, ומסומנת ב-ε.

שפה - קבוצה של מילים מעל א"ב נתון. מסומנת ב-L. לדוגמא: $L_1=\{\varepsilon,01,0110\}$. העוצמה של השפה בדוגמא היא $|L_1|=3$. שפה יכולה להיות ריקה $L_2=\{\emptyset\}$, והיא יכולה להיות גם אינסופית $L_3=\{0,00,000,...\}$ ואז העוצמה שלה היא: $|L_3|=\aleph_0$.

Σ^* - נסמן ב-Σ^* את קבוצת <u>כל</u> המילים מעל א"ב Σ. אם לדוגמא $\Sigma=\{0,1\}$ אזי $\Sigma^*=\{\varepsilon,0,1,00,01,10,11,...\}$.

עבור כל קבוצת סימנים א"ב Σ סופית עם n סימנים $n=|\Sigma|$, אזי לכל i טבעי ישנן n^i מילים באורך i, כלומר מילים בעלות i אותיות. העוצמה של כל Σ^* מעל קבוצה א"ב Σ כלשהי היא $|\Sigma^*|=\aleph_0$, כלומר קבוצה בת מנייה.

⊛ לכל שפה L מעל Σ מתקיים $L\subseteq\Sigma^*$. ומכאן נובע כי $|L|\leq\aleph_0$.

תכונות - נוכל להגדיר שפה L באמצעות תכונות מסוימות, וכך בדר"כ נעבוד. לדוגמא:

עבור $\Sigma=\{0,1\}$ נוכל להגדיר L כשפת כל המילים שמתחילות ב-0.

עבור $\Sigma=\{a,b\}$ נוכל להגדיר L כשפת כל המילים המכילות את המילה $w=aba$.

עבור $\Sigma=\{0,1,2...9\}$ נוכל להגדיר L כשפת כל המילים שמתחלקות ב-3.

⊛ נשים לב מהדוגמאות לעיל ש-ε לא בהכרח מוגדרת בכל שפה.

ב) פעולות על מילים

שרשור - של מילים w_1,w_2 היא המילה המתקבלת מכתיבת אותיות w_1 ולאחריה w_2. מסמנים כך $w_1\circ w_2$ (או w_1w_2). שרשור מילים אינה פעולה חילופית, יתכן $w_1\circ w_2\neq w_2\circ w_1$.

לדוגמא: $w_1=aab,\ w_2=abc \Rightarrow w_1\circ w_2=aababc \neq w_2\circ w_1=abcaab$

[illegible]רי: אוטומטים
[illegible]ות פורמליו[illegible]
[illegible]ניברסיטה
הפתוחה.

· [illegible] 3 מטלו[illegible]

[illegible] השפות מעל
[illegible] היא $|P(\Sigma^*)|$
[illegible]ין אינה בת
[illegible] לפי משפט קנטור

28/10 בס"ד

②

ניתן לשרשר מילה עם עצמה $w\circ w$. שרשור עם ε יתן לנו את המילה עצמה: $w\circ\varepsilon=\varepsilon\circ w=w$

ניתן גם לשרשר כמה מילים, זוהי פעולה קיבוצית כלומר אין חשיבות לסדר השרשור.

<u>חזקה</u> - היא שרשור של מילה עם עצמה i פעמים. $w^0=\varepsilon$, $w^i=\underbrace{w\circ w\circ w\circ\cdots\circ w}_{i\text{ פעמים}}$

הגדרה רקורסיבית של חזקה היא שאם $w^0=\varepsilon$ אזי לכל $i\ge 1$ נגדיר $w^i=w\circ w^{i-1}$.

הגדרה זו תועיל לנו כאשר נרצה להוכיח טענה באינדוקציה.

<u>היפוך</u> - תהי מילה w עם n אותיות $w=a_1a_2\dots a_n$ כך שלכל $i\in[n]$ $a_i\in\Sigma$, אזי ההיפוך של w, המסומן w^R (Reverse) היא: $w^R=a_n\dots a_2a_1$. אם $w=\varepsilon$ אזי $w^R=w=\varepsilon$.

לדוגמא: אם $w=aab$ אזי $w^R=baa$.

הגדרה רקורסיבית של היפוך: נסמן ב- $|w|$ את אורך המילה w, כלומר מספר האותיות ב- w. אם $|w|=0$ כלומר $w=\varepsilon$ אז $w^R=w$. אם $|w|>0$ כך ש- $w=a_1a_2\dots a_n$ כאשר $n\ge 1$, אזי $w^R=a_n\circ(a_1a_2\dots a_{n-1})^R$.

ג) <u>פעולות על שפות</u>

(1) <u>איחוד</u> - $L_1\cup L_2=\{w\mid w\in L_1 \cup w\in L_2\}$

(2) <u>חיתוך</u> - $L_1\cap L_2=\{w\mid w\in L_1 \cap w\in L_2\}$

(3) <u>הפרש</u> - $L_1\setminus L_2=\{w\mid w\in L_1 \cap w\notin L_2\}$

(4) <u>משלים</u> - $\overline{L}=\Sigma^*\setminus L$

דוגמא: עבור $\Sigma=\{0,1\}$ נגדיר L כשפת כל המילים שמתחילות ב-0: $L=\{w\mid\exists u\in\Sigma^*, w=0u\}$

המשלים של L הוא קבוצת כל המילים שלא מתחילות ב-0, או בצורה מתמטית:

$$\overline{L}=\{\forall u\in\Sigma^*,\ w\ne 0u\}=\{w\mid\exists u\in\Sigma^*,\ w=1u\}\cup\{\varepsilon\}$$

(5) <u>שרשור</u> - יהי שתי שפות L_1,L_2 אזי השרשור של L_1 עם L_2, המסומן $L_1\circ L_2$ (או L_1L_2) היא שפה שבה כל המילים מורכבות מכל האפשרויות לשרשר מילה מ-L_1 עם מילה מ-L_2.

ובשפה מתמטית: $L_1\circ L_2=\{w\mid\exists u\in L_1,\exists x\in L_2,\ w=u\circ x\}$.

דוגמא: $L_1=\{\varepsilon,01\}$, $L_2=\{01,1,10,\varepsilon\}$,

$L_1\circ L_2=\{\varepsilon,01,1,10,0101,011,0110\}$

מכיוון שאפשר לקבל מילים זהות מסך השרשורים תמיד מתקיים $|L_1\circ L_2|\le|L_1|\cdot|L_2|$

(6) חזקה - נסמן חזקה של שפה L^i. המשמעות היא לשרשר את שפה L עם עצמה i פעמים

כלומר $L^i = \underbrace{L\circ L\circ L\circ\ldots\circ L}_{i \text{ פעמים}}$. כאשר $i=0$ עבור כל שפה נקבל $L^0=\{\varepsilon\}$.

(7) איטרציה - היא פעולה מתמטית על שפה שמשמעותה היא איחוד אינסופי של כל החזקות האפשריות של השפה. מסמנים איטרציה של שפה L^*. מכיוון ש-$L^0=\{\varepsilon\}$, אזי תמיד $\varepsilon\in L^*$.

$$L^* = \bigcup_{i\in N} L^i = L^0\cup L^1\cup L^2\cup\ldots\cup L^\infty$$

⊛ כאשר $L=\{\varepsilon\}$ אזי כל חזקה גם היא ε לכן האיטרציה של L היא $L^*=\{\varepsilon\}$.

כאשר $L=\{\emptyset\}$ אזי $L^0=\{\varepsilon\}$ וכל שאר החזקות $\emptyset$ לכן סך האיחודים יתנו $\{\varepsilon\}=L^*$.

המסקנה היא שעבור כל שפה נקבל ש-L^* מכיל לפחות את ε, והיא אף פעם לא $\emptyset$.

⊛ כאשר $L=\Sigma$ קבוצת סימנים כלשהי, אזי $L^*=\Sigma^0\cup\Sigma^1\cup\Sigma^2\cup\ldots$, כלומר קבוצת כל המילים האפשריות מעל Σ שאותה סימנו Σ^*.

⊛ תהי w מילה מעל Σ. נסמן ב-$|w|$ את האורך של w, אזי עבור $|w|=i$ מתקיים $w\in\Sigma^i$, ומכאן שכל $w\in\Sigma^*$ לפי הגדרת האיטרציה.

משפט: אם בשפה L קיימת לפחות מילה אחת שאינה ε, כלומר קיים $w\in L$ כך ש-$|w|\ge 1$ אזי L^* היא בהכרח אינסופית ועוצמתה $|L^*|=\aleph_0$.

הוכחה: (1) לכל שפה L בהכרח $L^*\subseteq\Sigma^*$, ולכן $|L^*|\le\aleph_0$. הוכחנו בסעיף ב'.

(2) $w, w^2, w^3,\ldots$ היא קבוצה בת מנ״ה של מילים ששייכות ל-L^*, לכן $\aleph_0\le|L^*|$.

מ-(1)+(2) לפי משפט קנטור-שרדר-ברנשטיין נקבל $|L^*|=\aleph_0$.

שאלה: תהי $\Sigma=\{0,1\}$, ותהי L שפת כל המילים שמתחילות ב-0 מצא את L^*.

תשובה: $L^*=\{0\}\circ\Sigma^*\cup\{\varepsilon\}$. נראה הכלה דו כיוונית.

(1) $L^*\subseteq\{0\}\circ\Sigma^*\cup\{\varepsilon\}$ - תהי $w\in L^*$. אם $w=\varepsilon$ סיימנו לכן נניח $w\neq\varepsilon$.
אם כן קיים $0<i$ כך ש-$w\in L^i$. ומכאן שקיימות i מילים $u_1,u_2,\ldots,u_i\in L$
כך ש-$w=u_1\circ u_2\circ\ldots\circ u_i$. המילה הראשונה $u_1\in L$ ולכן מתחילה ב-0 ואזי גם
w מתחילה ב-0, כלומר $w\in\{0\}\cdot\Sigma^*$. ובפרט $w\in\{0\}\cdot\Sigma^*\cup\{\varepsilon\}$.

(2) $\{0\}\circ\Sigma^*\cup\{\varepsilon\}\subseteq L^*$ - תהי $w\in\{0\}\cdot\Sigma^*\cup\{\varepsilon\}$, אם $w=\varepsilon$ סיימנו, לכן נניח $w\neq\varepsilon$,
אזי $w\in\{0\}\cdot\Sigma^*$. ומכאן שקיים $u\in\Sigma^*$ כך ש-$w=0\cdot u$, ולכן $w\in L$. ומכיוון
ש-$L\subseteq L^*$ נקבל $w\in L^*$. מ.ש.ל.

בס"ד

4/11

(4)

<u>משפט:</u> מתקיים $(L^*)^* = L^*$.

הוכחה: נראה הכלה דו כיוונית

(1) מההגדרה של איטרציה ברור כי $L^* \subseteq (L^*)^*$.

(2) נוכיח ש-$(L^*)^* \subseteq L^*$. תהי $w \in (L^*)^*$, כך שקיים $i \in \mathbb{N}$ כך ש-$w \in (L^*)^i$.
לכן קיימות $u_1, u_2, \ldots, u_i \in L^*$ כך ש-$w = u_1 \circ u_2 \circ \ldots \circ u_i$. עבור כל $1 \le j \le i$
נפרק כל u_j לשרשור של l_j מילים, שהרי כל $u_j \in L^*$. l_j הוא מספר
המילים שנמצאות בפירוק של המילה u_j, כאשר $l_j \in \mathbb{N}$, כלומר קיימים
$x_{j1}, x_{j2}, \ldots, x_{jl_j} \in L$ כך ש-$u_j = x_{j1} \circ x_{j2} \circ \ldots \circ x_{jl_j}$ עבור כל $1 \le j \le i$. נגדיר את w
כך:

$$w = \underbrace{x_{11} \cdot x_{12} \cdot \ldots \cdot x_{1l_1}}_{u_1} \cdot \underbrace{x_{21} \cdot x_{22} \cdot \ldots \cdot x_{2l_2}}_{u_2} \cdot \ldots \ldots \cdot \underbrace{x_{i1} \cdot x_{i2} \cdot \ldots \cdot x_{il_i}}_{u_i}$$

(8) <u>היפוך של שפות</u> - תהי L שפה כלשהי, אזי ההיפוך של L, המסומן L^R, היא שפה
שבה לכל המילים אנו עושים היפוך. $L^R = \{w^R \mid w \in L\}$.

משפט: מתקיים $(L^R)^R = L$. פשוט ואין צורך בהוכחה.