לקראת המבחן

הגדרות:

- א"ב - $\sum$ קב' סופית לא ריקה של אותיות
- מילה = סדרה סופית מעל א"ב $\sum$
- שפה: קבוצה של מילים, מעל א"ב $\sum$, נסמן ב L
- מילה ריקה: ε = מילה ללא אותיות
- שפה ריקה : $L=\emptyset$, נשים לב ש $L=\emptyset\neq\{\varepsilon\}$
- אורך של מילה w , $|w|$ = מס' האותיות ב w
- $\sum^*$ = אוסף כל המילים מעל א"ב $\sum$,
  - $\varepsilon\in\sum^*$
  - כל L מעל $\sum$ מקיימת $L\subseteq\sum^*$
- $\left|\sum^*\right|=\aleph_0$ קב' של סדרות סופיות מעל קב' סימנים סופית (לא ריקה).

פעולות על מילים

- שרשור של מילים בהנתן $w_1,w_2\in L$ המילה $w_1\cdot w_2$ היא המילה המתקבלת מכתיבת אותיות w_2 אחרי w_1
  - שרשור מילים אינו פעולה חילופית, כלומר $w_1\cdot w_2\neq w_2\cdot w_1$
  - שרשור היא פעולה קיבוצית, כלומר : $(w_1\cdot w_2)\cdot w_3=w_1\cdot(w_2\cdot w_3)$
  - שרשור של מילים לא בהכרח משמר סגירות בשפה, כלומר יתכן $w_1\cdot w_2\notin L$
- חזקה של מילים:
  - הגדרה פשוטה: $i\in\mathbb{N}$, w^i: $w^i=w\cdot w\cdot w\cdot w...$
  - הגדרה רקורסיבית: $$\begin{gathered}w^0=\varepsilon\\ w^1=w\\ w^i=w\cdot w^{i-1}\end{gathered}$$
- היפוך $(reverse)$: w^R
  - הגדרה פשוטה: $$\begin{gathered}w=a_1,a_1,...,a_n\\ w^R=a_n,a_{n-1},....a_1\end{gathered}$$
  - הגדרה רקורסיבית: $$\begin{gathered}w=\overbrace{a_1,a_2,...}^{t}a_n\\ w^R=a_nt^R\end{gathered}$$. במקרה ו $w=\varepsilon$ אז $w^R=w$

פעולות על שפות

תהיינה L_1,L_2 שפות מעל א"ב $\sum$ אז :

- איחוד - $L_1\cup L_2=\{w|w\in L_1\vee w\in L_2\}$
- חיתוך - $L_1\cap L_2=\{w|w\in L_1\wedge w\in L_2\}$

- <u>משלים</u> - $L_1 = \bar{\sum}^* \backslash L_1$ וכן $L_1 \backslash L_2 = \{w | w \in L_1 \wedge w \notin L_2\}$
- <u>שרשור שפות</u> $L_1 \cdot L_2 = \{w | \exists u \in L_1, \exists v \in L_2, w = u \cdot v\}$
- <u>איטרציה</u> : $L^* = \bigcup_{i \in \mathbb{N}} L^i = L^0 \cup L^1 \cup ...$
- <u>היפוך</u>: - L^R , הגדרה: $L^R = \{w | \, w^R \in L\}$.

אוטומט סופי דטרמינסטי

- נאמר שהאוטומט **מקבל** קלט w אם הוא מסיים את קריאת w במצב מקבל .
  - אחרת נאמר שהוא **דוחה** את w.
- אוטומט סופי דטרמינסטי A נתון על ידי $A = (\sum_A, Q_A, q_{0A}, F, \delta_A)$, כאשר
  - <u>א"ב</u> – כל אותיות הקלט האפשריות עבור האוטומט. מספר האותיות בא"ב זה חייב להיות סופי וגדול מ־0.
    - נסמן ב $\sum_A$
  - <u>מצבים</u> – כל המצבים שבהם יכול האוטומט להימצא. מספר המצבים חייב להיות סופי וגדול מ־0.
    - נסמן ב Q_A - קבוצה סופית לא ריקה של מצבים
  - <u>מצב התחלתי</u> – המצב שממנו מתחיל האוטומט את מסלול החישוב על כל מילת קלט . קבוצת מצבים מקבלים – קבוצה מתוך קבוצת המצבים, המכילה 0 מצבים או יותר .
    - q_{0A} - מצב התחלתי $q_{0A} \in Q_A$
    - F_A - קבוצת מצבים מקבלים $F_A \subseteq Q_A$
  - <u>פונקציית מעברים</u> – לכל זוג של מצב ואות, פונקציה זו מתאימה מצב (אחד ויחיד) שאליו עובר האוטומט כאשר במצב זה נקראת אות זו.
    - F_A - קבוצת מצבים מקבלים $F_A \subseteq Q_A$
- <u>פונקציית המעבר</u>:
  - <u>בסיס</u> $\delta(q, \varepsilon) = q$
  - <u>צעד</u>: $\delta(q, wa) = \delta(\delta(q, w), a)$ כאשר w היא מחרוזת ו - a היא אות.
- עבור מצב q באוטומט א"ד נאמר ש q הוא **מצב בור** אם לכל $a \in \sum$ מתקיים $\delta(q, a) = \emptyset$.

שפה של אוטומט דטרמינסטי:

- אם $A = (\sum_A, Q_A, q_{0A}, F_A, \delta_A)$הוא אוטומט אז $L(A)$ היא השפה שמקבל האוטומט. = אוסף כל המחרוזות w כך ש $\delta(q_0, w)$ שייך ל F . כלומר:$L(A) = \{w \in \sum^* | (\delta_0, w) \in F\}$

  כלומר אוסף כל המחרוזות שמקבל האוטומט - אוסף כל המחרוזות המהוות מסלול ממצב ההתחלה למצב מקבל.
- לכל $q \in Q$ שפת המצב $L_A(q)$ היא השפה הבאה: $L_A(q) = \{w \in \sum^* | \delta(q_0 = w)\}$

  מתקיים $L(A) = \bigcup_{q \in F} L_A(q)$
- <u>הגדרה</u>: שפה L היא רגולרית אם היא מתקבלת ע"י אוטומט סופי דטרמיניסטי.

אוטומט סופי לא דטרמינסטי

$A=(Q,\sum,q_0,\delta,F)$ כאשר:

- Q קבוצת מצבים, $\sum$ א"ב נתון, <u>פונקציית מעברים δ</u> , $q_0\in Q$ מצב התחלתי, קבוצת מצבים מקבלים $F\subseteq Q$
- פונקציית המעברים δ: $\delta\left(q,a\right)$ היא **קבוצת** מצבים - הרחבה למחרוזת:
  - בסיס: $\delta\left(q,\varepsilon\right)=\left\{ q\right\}$
  - צעד: $\delta\left(q,wa\right)=\bigcup\limits_{p\in\delta(q,w)}\delta\left(p,a\right)$, סימון - להרחבה לקבוצת מצבים: $\delta\left(\boldsymbol{P},w\right)=\bigcup\limits_{\boldsymbol{q\in P}}\delta\left(q,w\right)$
- השפה של האוטומט היא קבוצת המחרוזות שהוא מקבל:

  כלומר $L(A)=\left\{ w\in\sum^{*}\left|\delta\left(q_{0},w\right)\cap F\neq\emptyset\right.\right\}$

אסל"ד עם מסעי ε

- אסל"ד עם מסעי ε - מסעי ε מאפשרים מעבר ממצב למצב על קלט ε
  - $\delta:Q\times\sum\cup\left\{ \varepsilon\right\} \rightarrow P(Q)$
- סגור של קבוצת מצבים:
  - סגור של מצב q = קבוצת המצבים שניתן להגיע אליהם מ q ע"י שימוש במסעי ε בלבד נסמן $CL(q)$
  - הסגורים של קבוצת מצבים P = איחוד כל הסגורים $CL(q)$ כאשר q שייך ל P
    - <u>בסיס</u>: $\delta'\left(q,\varepsilon\right)=CL(q)$
    - <u>צעד</u>: $\delta'\left(q,xa\right)=\bigcup\limits_{p\in\delta'(q,\varepsilon)}CL(\delta\left(p,\varepsilon\right)$

הרחבת מילים של פונקציות מעברים:

- באינדוקציה:
  - <u>בסיס</u>: $\delta'=\left(q,\varepsilon\right)=CL(q)$
  - <u>צעד</u>: $\delta'\left(q,xa\right)=\bigcup\limits_{p\in(q,x)}CL\left(\delta\left(p,a\right)\right)$

<u>הרעיון</u>: $\delta'\left(q,w\right)$ היא קבוצת המצבים שאליהם ניתן להגיע מ q ע"י קריאת w ושימוש אפשרי במסעי ε בכל שלב.

- $\sum^{+}$ כל המילים בשפה המכילות לפחות אות אחת (המילה הריקה בחוץ)

  <u>שפה של אוטומט לא דטרמיסנטי</u> עם מסעי ε היא קבוצת w עבורן $\delta'(q_0,w)$ מכילה מצב מקבל.

ביטויים רגולרים

- <u>הגדרה</u>: אוסף הביטויים הרגולרים מעל א"ב $\sum$ המסומן ב $R_{\sum}$ מוגדר **באינדוקצייה מבנית** באופן הבא:

  <u>אטומים</u>:
  - $\phi,\varepsilon\in R$
  - $\forall\sigma\in\sum,\sigma\in R$

  <u>פעולות יצירה</u>:
  - אם $r_1,r_2\in R$ אז $(r_1+r_2)\in R$, $(r_1\cdot r_2)\in R$

- אם $r\in R$ אז $(r^*)\in R$

<u>הגדרת שפה:</u> נגדיר את הפונקציה L מ R ל $2^{\sum^*}$:

- תהי $L[r]$ השפה שמציין הביטוי r:
  - $L\left[\phi\right]=\phi$
  - $L\left[\varepsilon\right]=\left\{ \varepsilon\right\}$
- לכל $\sigma\in\sum$ מתקיים $L\left[\sigma\right]=\left\{ \sigma\right\}$
- אם $r_1,r_2\in R$ אז:
  - $L\left[\left(r_{1}+r_{2}\right)\right]=L\left[r_{1}\right]+L\left[r_{2}\right]$
  - $L\left[\left(r_{1}\cdot r_{2}\right)\right]=L\left[r_{1}\right]\cdot L\left[r_{2}\right]$
- אם $r\in R$ אז $L\left[\left(r^{*}\right)\right]=\left(L\left[r\right]\right)^{*}$
- אם r ביטוי רגולרי נסמן ב $\left(r^{+}\right)$ את הביטוי $r\cdot\left(r^{*}\right)$
- סדר קדימויות בהשמטת סוגריים:
  - $*$
  - $\cdot$
  - $+$
  - (לרוב נשמיט את אופרטור השרשור)

משפטים

- פעולת שרשור על שפות **אינה חלופית,** יתכן - $L_{1}\cdot L_{2}\neq L_{2}\cdot L_{1}$
- פעולת שרשור על שפות היא **כן קיבוצית** $\left(L_{1}L_{2}\right)L_{3}=L_{1}\left(L_{2}L_{3}\right)$
- <u>טענה: $\left(L^{*}\right)^{*}=L^{*}$</u>- מסקנה מידית : אין טעם לבצע איטרציה יותר מפעם אחת.
- טענה: מתקיים $\delta\left(q,w_{1}\cdot w_{2}\right)=\delta\left(\delta\left(q,w_{1}\right),w_{2}\right)$
- תהיה קבוצת השפות מעל $\sum$, אז | = $\left|P\left(\sum^{*}\right)\right|=2^{\aleph_{0}}$ (מסקנה מיידית: רוב השפות אינן רגולריות)
- $L=\left\{ 0^{n}1^{n}|\geq1\right\}$ אינה שפה רגולרית
- אס"ד שקול לאסל"ד.
- אסל"ד ללא מסעי ε $\Longleftrightarrow$ אסלד עם מסעי ε .
- **סגירות השפות הרגולריות תחת פעולות בוליאניות**
  - **הכלה** - יהיו L_1,L_2, שפות כך ש: $L_{1}\subseteq L_{2}$
    - אם L_2 <u>רגולרית</u> אז L_1 <u>אינה בהכרח רגולרית</u> - ד"נ: $L_{1}=\left\{ a^{n}b^{n}|n>0\right\} \quad L_{2}=\left\{ \sum^{*}\text{where }\sum=\left\{ a,b\right\} \right\}$
    - אם L_1 <u>רגולרית</u> אז L_2 <u>אינה בהכרח רגורלית</u> - ד"נ: $L_{1}=\left\{ 01\right\} \quad L_{2}=\left\{ a^{n}b^{n}|n>0\right\}$
  - **משלים** - אם L רגולריות אז $\bar{L}$ רוגולרית.
  - **חיתוך** - אם L_1,L_2 רגולריות אז $L_{1}\cap L_{2}$ רגולרית

- **איחוד** - אם L_1, L_2 רגולרית אז $L_1 \cup L_2$ רגולרית
- **שרשור**- אם L_1, L_2 רגולריות אז $L_1 \cdot L_2$ רגולרית.
- **איטרציה** - אם L_1 רגולרית אז L_1^* רגולרית.
- **היפוך** - אם L_1 רגולרית אז L_1^R

- **למת הניפוח לשפות רגולריות**

תהי L שפה רגולרית. אז קיים n ב N כך שלכל מילה ב L שאורכה לפחות n קיים פירוק מהצורה $Z = uvw$ כאשר:

- $|uv| \leq n$
- $1 \leq |v|$
- $uv^iw \in L$ לכל $0 \leq i$

- **שלילת למת הניפוח**

נניח בשלילה ש L רגולרית **יהי** n מס' המובטח מלמת הניפות. נראה ש**קיימת** מילה $z \in L$ המקיימת $n \geq |z|$ ש**לכל** פירוק $z = uvw$, יסתור את התנאים בלמת הניפוח:

1. $(uv > n)|uv| \leq n$
2. $(1 > |v|)\ 1 \leq |v|$
3. $uv^iw \notin L$ **קיים** $0 \leq i$

- <u>שפות לא רגולריות - דוגמאות:</u>

- $L = \{0^n1^n | n \in \mathbb{N}\}$, אינה רגולרית
- $L = \left\{x \in \{a,b\}^* | \#_a(x) = \#_b(x)\right\}$ אינה רגולרית. (# - "כמות ה")
- $L = \left\{xx | x \in \{a,b\}^*\right\}$ אינה רולגרית.
- $L = \left\{a^{k^2} | k \in \mathbb{N}\right\}$ אינה רגולרית.

- לכל $r \in R$ מעל $\sum$ $L[r]$ היא שפה רגולרית, ($L[r] \iff$ קיום אוטמט)
- לכל שפה רגולרית $L \subseteq \sum^*$ קיים ביטוי רגולרי כך ש $L[r] = L$
- **משפט קליני** - משפחת השפות הרגולריות היא הקבוצה הקטנה ביותר המכילה את כל השפות הסופיות והסגורה תחת הפעולות הרגולריות