הפקולטה למדעי המחשב

אוטומטים ושפות פורמליות (236353)

מושגים בסיסיים

הרצאה 1

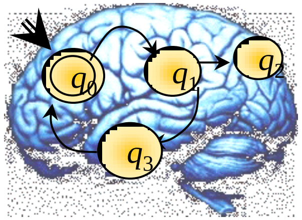
המצגת מבוססת על ספרם של פרופ' נסים פרנסיז ופרופ' שמואל זקס, "אוטומטים ושפות פורמליות", האוניברסיטה הפתוחה, 1987. גרסה ראשונה של מצגת זו הוכנה על-ידי אריאל ירושביץ, תחת הנחייתו של פרופ' שמואל זקס. בגרסה הנוכחית הוכנסו מספר שינויים ועדכונים על-ידי פרופ' יובל ישי, פרופ' מיכאל קמינסקי ופרופ' שמואל זקס.

אוטומטים ושפות פורמליות - מבוא

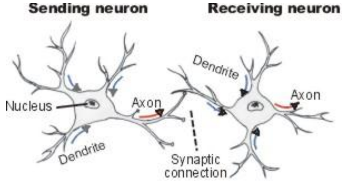
- קורס ראשון ב***מדעי*** המחשב. למה?
  - שאר הקורסים עד כה היו במתמטיקה או בתכנות (בהכללה גסה)
- קורס תיאורטי – מתמטי
- מהווה בסיס לתחומים רבים מאוד במדעי המחשב, תיאורטיים ומעשיים כאחד (ראה המשך)

אוטומטים ושפות פורמליות – מאין ולאן?
(סקירה לא מחייבת)

פעולת המוח



רשתות נוירונים



בלשנות (חישובית)

Sentence

Noun Phrase — *Verb Phrase*

Article + Adjective — *Noun Phrase* — *Verb* — *Adverb*

The little **boy** **went** **slowly**

תורת המיתוג

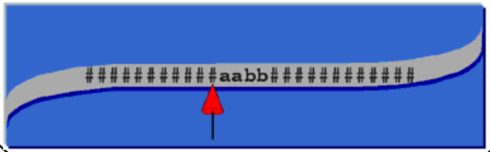
אוטומטים ושפות פורמליות

Shall I compare thee to a summer's day?

חישוביות

Q: "Halt?"

A: "@#\$%!!!"



קומפילציה

```
gcc -o MyKillerApp.exe
```

עיבוד שפות טבעיות

על מה הקורס? (הקדמה לא פורמלית)

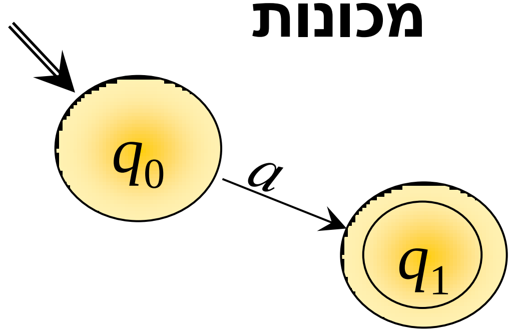
שפות

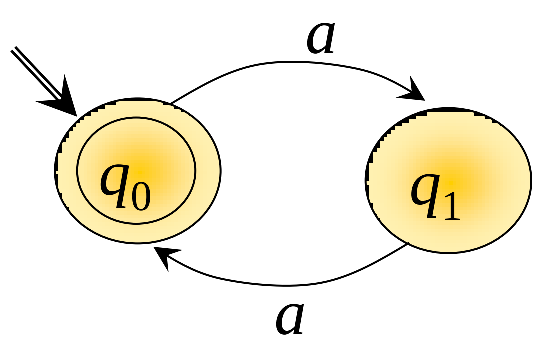
$\{a\}$

$\{w \mid |w| \equiv 0 \pmod 2\}$

$|x|$ מסמן את אורך המילה x

מכונות





דקדוקים

$S \rightarrow a$

ε היא המילה הריקה

$S \rightarrow \varepsilon$

$S \rightarrow Saa$

הגדרות

הגדרה

אלף בית Σ הינו קבוצה **סופית** כלשהי.

דוגמאות

- $\Sigma_1 = \{a,b,c\}$
- $\Sigma_2 = \{0,1\}$
- $\Sigma_3 = \{(x,y) \mid x \in \{0,1\}, y \in \{a,b,c\}\}$

☝מוסכמות סימון:

- נשתמש בדרך כלל באותיות יווניות גדולות כגון Σ, Γ, Δ לציון א"ב
- נשתמש באותיות יווניות קטנות כגון σ, γ לציון *משתני* אותיות

הגדרה

אות $\sigma \in \Sigma$ הינה איבר מתוך א"ב נתון.

מילים

הגדרה

מילה **(word)** w מעל א"ב Σ הינה סידרה **סופית** של אותיות.

דוגמאות

- aaa היא מילה מעל א"ב $\{a\}$
- 1313d3$$# היא מילה מעל {1,3,d,$,#,%,&}
- $(0,b)(0,a)(1,c)$ היא מילה מעל $\{(x,y) \mid x\in\{0,1\}, y\in\{a,b,c\}\}$

☝מוסכמות סימון

- בהקשרים אחרים נהוג לסמן סדרות כך: a,b,c,d,e. כאן אנו משמיטים את הפסיקים ורושמים $abcde$
- נשתמש בדרך כלל באותיות לטיניות מסוף הא"ב כגון w,v,x,y,z לציון משתני מילים

עוד על מילים

- תהי $w = s_1 s_2 \ldots s_n$. נסמן את אורכה ב $|w|$.
כלומר $|w| = n$
- נסמן ב ε את המילה הריקה: $|\varepsilon|=0$
- יהיו $w_1 = \sigma_1 \ldots \sigma_n , w_2 = \gamma_1 \ldots \gamma_m$ מילים. ***שרשור*** **(*concatenation*)** של שתי המילים יסומן ב $w_1 \cdot w_2$ ויוגדר כ"הדבקתן" זו לזו עם חשיבות לסדר, לכדי יצירת מילה חדשה:
- $w_1 \cdot w_2 = \sigma_1 \ldots \sigma_n \cdot \gamma_1 \ldots \gamma_m = \sigma_1 \ldots \sigma_n \gamma_1 \ldots \gamma_m$
  - בד"כ נשמיט את אופרטור השרשור
  - נשים לב כי: $\forall w,\ w \cdot \varepsilon = \varepsilon w = w$

לא הגדרנו מעל איזה א"ב כל מילה.
האם שתיהן מעל אותו א"ב?
האם זה משנה??

עוד על מילים (המשך)

- תכונות השרשור
  - $(w_1 \cdot w_2) \cdot w_3 = w_1 \cdot (w_2 \cdot w_3)$
  - $|w_1 \cdot w_2| = |w_1| + |w_2|$
- איך מוכיחים את התכונות הנ"ל?

הגדרה

u היא ***רישא*** של w אמ"מ $\exists v, u \cdot v = w$

u היא ***סיפא*** של w אמ"מ $\exists v, v \cdot u = w$

- נאמר ש u היא רישא (סיפא) טריוויאלית של w אם $u = w$ או $u = \varepsilon$.
- נאמר שרישא (סיפא) היא רישא (סיפא) ממש, או אמיתית, אם היא אינה טריוויאלית.

עוד על מילים (המשך)

- **הגדרה**

חזקה של מילה w

- $w^0 = \varepsilon$ ו-
- $w^{i+1} = w^i \cdot w$

- כלומר,

$$w^i = \underbrace{ww\ldots w}_{i}$$

- נסמן את מספר המופעים של האות σ במילה w ב- $\#_\sigma(w)$

לדוגמא: $\#_a(aaba) = 3$,$\#_b(aaba) = 1$,$\#_c(aaba) = 0$

שפות (פורמליות)

- **הגדרה**

  שפה (פורמלית) L מעל א"ב Σ היא קבוצת מילים מעל אותו א"ב.

- דוגמאות לשפות מעל $\Sigma = \{a,b\}$:
  - $\{aab,b\}$
  - $\{aaaa,\varepsilon\}$
  - $\{w | \#_a(w) = \#_b(w)\}$
  - $\{\varepsilon\}$
  - $\varnothing$
  - $\{w | \ \#_a(w) \mid \#_b(w)\}$
  - $\{w | \ |w| \ is\ prime\}$
  - $\{a^n b^n | n \in \mathbb{N}\}$

פעולות על שפות

- **הגדרה**

שרשור של שפות L_2, L_1 הוא השפה הבאה:

$$L_1 \cdot L_2 = \{w \cdot v | w \in L_1 \wedge v \in L_2\}$$

- **הגדרה**

חזקה של שפה L

- $L^0 = \{\varepsilon\}$ ו-
- $L^{i+1} = L^i \cdot L$

- סגור קליני: $L^* = \bigcup_{i=0}^{\infty} L^i$
- הסגור החיובי: $L^+ = \bigcup_{i=1}^{\infty} L^i$
- בהנתן א"ב Σ מסמנים את קבוצת כל המילים מעל א"ב זה ב- Σ^*

שפות (המשך)

- נשים לב כי לכל L, $L^* \in \varepsilon$.

- מהו איבר היחידה ביחס לשרשור שפות?
- למה שווה $L\cdot\varnothing$?
- האם $(L_1 \cup L_2)L_3 = L_1L_3 \cup L_2L_3$?
- האם $(L_1 \cap L_2)L_3 = L_1L_3 \cap L_2L_3$?
- האם $(L_1L_2)^* = L_1^*L_2^*$?

סיכום ודגשים

- הגדרנו את המושגים א"ב, מלים ושפות
- הגדרנו פעולות על המושגים הנ"ל:
  - אורך (למילים)
  - שרשור (למילים ושפות)
  - חזקה (למילים ושפות)
  - רישא/סיפא (למילים)
- נשים לב:
  - א"ב הוא תמיד קבוצה סופית
  - אורך של מילה הוא תמיד סופי
  - שפות יכולות להיות סופיות או אינסופיות
    - עוד על עוצמות הקבוצות שהכרנו - בתרגול

ספרות מומלצת

- ש. זקס, נ. פרנסיז, "*אוטומטים ושפות ופרמליות*", הוצאת האוניברסיטה הפתוחה, 2000.
- Hopcroft & Ullman, "*Introduction to Automata Theory, Languages, and Computation*", Addison-Wesley, 1979
(במהדורה החדשה שונו מספר הוכחות.)