

## הרצאה 6

אלגוריתמים חמדניים, עצי הופמן

## הקדמה

רוצים לשמור קובץ טקסט על הדיסק בצורה חסכונית. אפשרות אחת היא לקודד כל תו בטקסט במספר סיביות קבוע. מספר הסיביות שנזדקק לכל תו הוא  $\lceil \log |\Sigma| \rceil$ . אפשרות נוספת היא לקודד כל תו במספר סיביות שונה. נשים לב שקידוד כזה יכול להיות חסכוני יותר כאשר יש שוני בין שכיחויות התווים בטקסט.

**דוגמה:**

עבור הא"ב  $\{A, B, C, D\}$  והמחרוזת הבאה: AAABCD קידוד באורך קבוע יהיה באורך  $6 \times 2 = 12$ .

A	1
B	01
C	001
D	000

לעומת זאת, אם נבחר את הקידוד הבא

אז אורך הקידוד יהיה 11 בלבד.

**הגדרה 1** (קוד בינרי). בהינתן א"ב סופי  $\Sigma$  קידוד הוא פונקציה שמעפה כל תו בא"ב למחרוזת בינרית  $c: \Sigma \rightarrow \{0, 1\}^*$

**הגדרה 2** (הרחבה של קוד). הרחבה של קוד היא פונקציה  $c: \Sigma^* \rightarrow \{0, 1\}^*$  שמוגדרת להיות  $c(t_1 \dots t_k) = c(t_1) \dots c(t_k)$

## תכונות

נבחן שלושה קידודים שונים לא"ב  $\{A, B, C, D\}$

$$c_1 = \begin{array}{|c|c|} \hline A & 1 \\ \hline B & 01 \\ \hline C & 001 \\ \hline D & 000 \\ \hline \end{array} \quad c_2 = \begin{array}{|c|c|} \hline A & 0 \\ \hline B & 01 \\ \hline C & 011 \\ \hline D & 111 \\ \hline \end{array} \quad c_3 = \begin{array}{|c|c|} \hline A & 1 \\ \hline B & 01 \\ \hline C & 011 \\ \hline D & 111 \\ \hline \end{array}$$

באופן טבעי נדרוש שהקוד יהיה ניתן לפענוח (חד פעמי), כלומר נרצה שההרחבה תהיה פונקציה חד חד ערכית.

**דוגמה:** ניתן לפענח את  $c_1$ , ו- $c_2$ , אבל לא את  $c_3$ .

תכונה רצויה היא שנוכל לפענח כל תו ברגע שקראנו את המילה שמקודדת אותו (פענוח מידי).

**דוגמה:** התכונה מתקיימת עבור  $c_1$ , אבל לא מתקיימת עבור  $c_2$ , ו- $c_3$ .

## קודים חסרי רישות

קוד  $c$  יקרא חסר רישות אם לא קיימים  $a, b \in \Sigma$  כך ש- $c(a)$  רישא של  $c(b)$   
 קל לראות שקודים חסרי רישות ניתנים לפענוח וכן לפענוח מידי. מעבר לכך המשפט הבא (ללא הוכחה) מראה שלמטרתנו מספיק להתמקד בקודים חסרי רישות.

**משפט 1.** לכל קוד חד פעמי  $c$  קיים קוד חסר רישות  $c'$  כך שלכל  $a \in \Sigma$  מתקיים  $|c(a)| = |c'(a)|$ .

## קוד חסר רישות כעץ בינרי

ניתן לייצג כל קוד חסר רישות כעץ בינרי, למשל את הקוד  $c_1$  ניתן לייצג על ידי העץ הבא:

