

דחיסת נתונים - מטלה 3

שאלה 1

נשים לב כי $\text{firstcode}[l+1]$ היא מילת הקוד הראשונה באורך $l+1$,

והערך $\text{num}[l+1]$ הוא מספר מילות הקוד באורך $l+1$.

העץ הקנוני המתקבל לאחר בניה זו הינו "משוך שמאלה" כפי שלמדנו בשיעור.

טענת עזר:

- נזכיר כי בבניית עץ הופמן הקנוני לאחר הבניה מתקבלים אורכי קוד בסדר יורד. ועבור אורך קוד כלשהו נקבל מילות קוד שהן מספרים בינאריים עוקבים. נניח אב (z) לשני בנים (y) , (x) ונניח כי (x) הוא השמאלי מבניהם. מבניית העץ הקנוני מתקיים שאורך הקוד המינימלי בתת עץ (x) הוא לכל הפחות האורך המקסימלי בתת עץ (y) [תרשים 1 למטה]. מבניה זו ומהעקביות, נובע כי $\text{firstcode}[l+1]$ היא המילה השמאלית ביותר בעץ באורך $l+1$. בנוסף, המילה הימנית ביותר עבור אורך מסוים מסתיימת בעלה ימני. שאם לא כן, ע"פ הטענה נקבל אב עם בן שמאלי יחיד ואין זה יכול לקרות מהגדרת עץ הופמן [תרשים 2 למטה].

נניח בשלילה שהביטוי $\text{firstcode}[l+1] + \text{num}[l+1]$ הוא אי-זוגי. כלומר, אחד מהביטויים הוא זוגי ואחד אי-זוגי.

אם $\text{num}[l+1]$ אי-זוגי, אזי מטענת העזר נובע כי המילה השמאלית ביותר באורך $l+1$ היא בן ימני יחיד לאב כלשהו באורך l [תרשים 3 למטה] והבן השמאלי לאב זה הוא תת עץ עבור מילים ארוכות יותר ולכן מסתיימת בתו 1. כלומר $\text{firstcode}[l+1]$ מהצורה $X1$ ולכן ערכה הוא אי-זוגי – סתירה.

אם $\text{num}[l+1]$ זוגי, אזי מטענת העזר נובע כי המילה השמאלית ביותר באורך $l+1$ היא בן שמאלי לאב באורך l ולכן מסתיימת בתו 0. כלומר $\text{firstcode}[l+1]$ מהצורה $X0$ ולכן ערכה הוא זוגי – סתירה.

לכן, נסיק כי הביטוי $\text{firstcode}[l+1] + \text{num}[l+1]$ הינו זוגי.

תרשים 3	תרשים 2	תרשים 1

שאלה 2

נדחס את ההודעה abracadabra באמצעות קוד הופמן דינאמי.

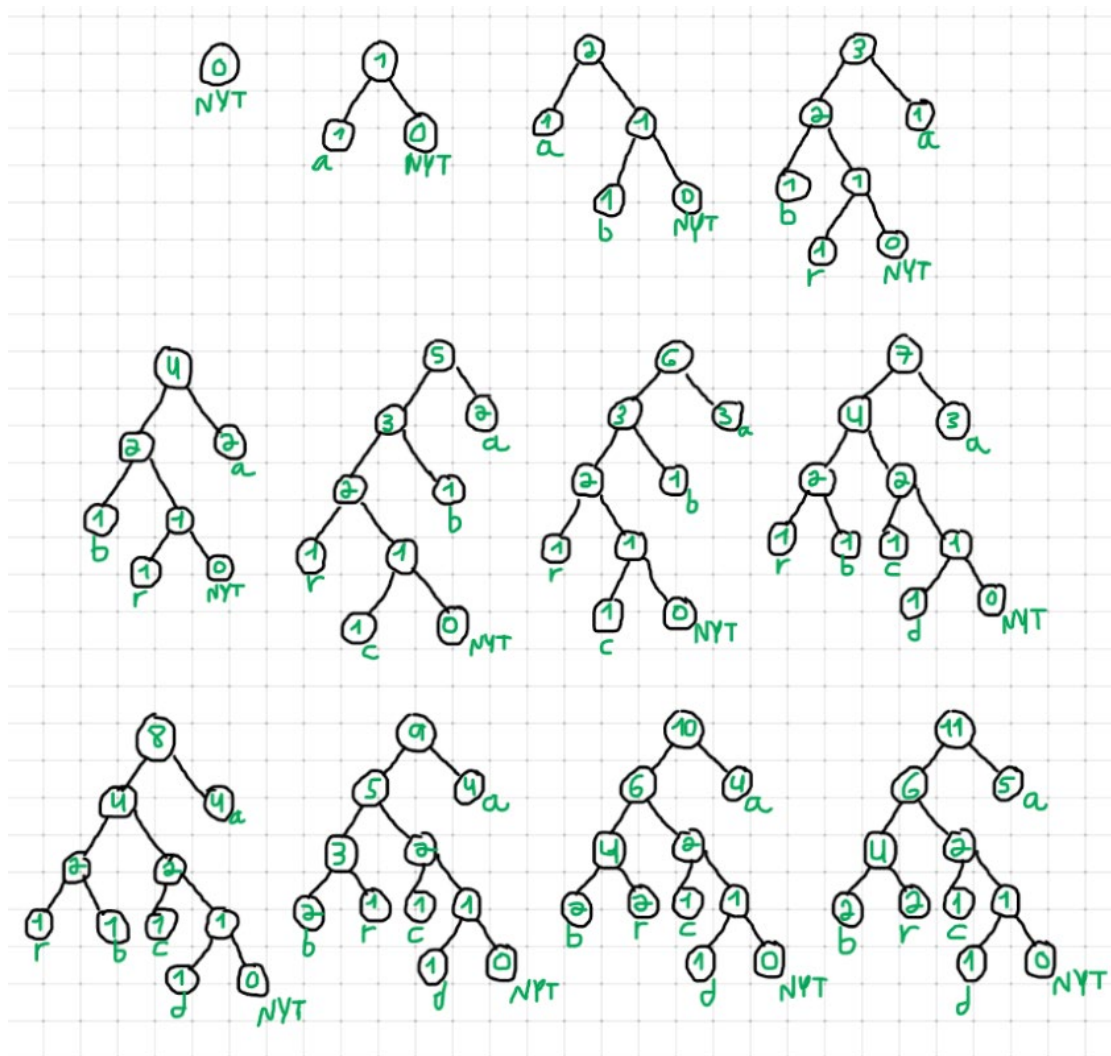
ההודעה הדחוסה הינה: $NYTaNYTbNYTr1NYTc1NYTd10010011=a1b11r1011c10011d10010011$

כאשר NYT הוא המסלול אל NYT בשלב המתאים וכל אות בהודעה מיוצגת ע"י קוד האסקי שלה (8 סיביות).

הטבלה הסופית הינה:

מילת קוד	תו	מילת קוד	תו	מילת קוד	תו
1	a	010	c	0110	d
000	b	001	r		

עץ ההופמן בכל שלב (משמאל לימין):



ב.

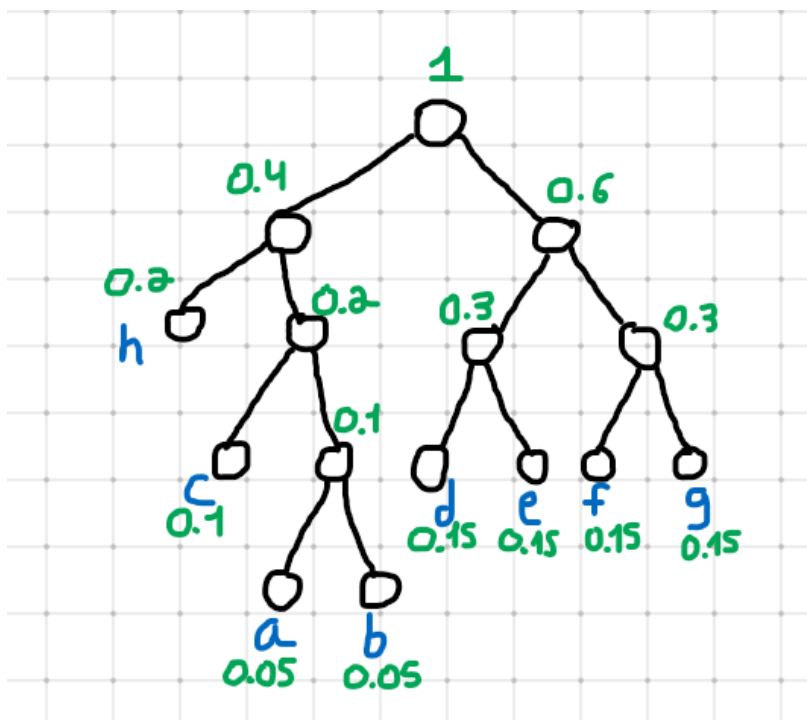
עבור קוד הופמן דינאמי: קיבלנו הודעה באורך $5 \cdot 8 + 20 = 60$ סיביות. (20 סיביות לניתוב בעץ + הצגת 5 תווים באסקי ע"י 8 סיביות).

עבור מילות קוד באורך זהה (עם הידר): נקבל $3 \cdot 11 + 5 \cdot 3 + 5 \cdot 8 = 88$ סיביות (היידר המורכב מהצגת כל תו ע"י 3 סיביות והערך המתאים לו בטבלת האסקי [3 סיביות מספיקות בשביל להציג 5 תווים] + אורך ההודעה*אורך מילת קוד).

עבור מילות קוד באורך זהה (ללא הידר): נקבל $3 \cdot 11 = 33$ סיביות (אורך ההודעה*אורך מילת קוד).

שאלה 3

נניח את עץ הופמן הבא:



כאשר הסתברויות הופעת האותיות הן בצבע ירוק.

(ניתן לראות שע"פ חלוקת ההסתברויות העץ יכול להתקבל מאלגוריתם הופמן).

מילות הקוד הינם:

מילת קוד	תו	מילת קוד	תו
0110	a	101	e
0111	b	110	f
010	c	111	g
100	d	11	h

נפעיל על העץ את האלגוריתם שלמדנו בשיעור לקבלת עץ קנוני.

$\text{num}[4]=2, \text{num}[3]=5, \text{num}[2]=1$

מילת הקוד הראשונה באורך 4 הינה 0000 ע"פ שורה 6 באלגוריתם. 0 ב-4 סיביות.

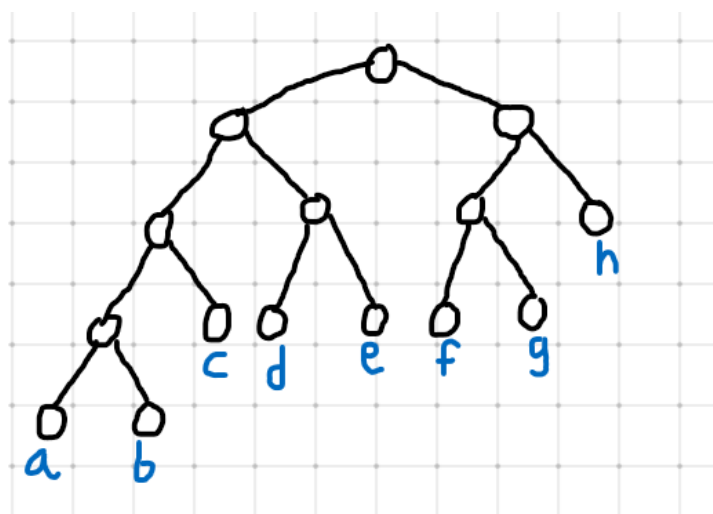
מילת הקוד הראשונה באורך 3 הינה 001 ע"פ שורה 7 באלגוריתם. $(0+2)/2=1$ ב-3 סיביות.

מילת הקוד הראשונה באורך 2 הינה 11 ע"פ שורה 7 באלגוריתם. $(1+5)/2=3$ ב-2 סיביות.

מכאן על פי הלולאה השנייה והשלישית באלגוריתם נקבל:

מילת קוד	תו	מילת קוד	תו
0000	a	011	e
0001	b	100	f
001	c	101	g
010	d	11	h

כלומר נקבל את העץ:



כאשר העץ קנוני ומשוך שמאלה (ושקול לעץ הקנוני המשוך ימינה).

כעת, נבחין בעץ השלד הנוצר מעץ ההופמן שאינו קנוני לעץ השלד הנוצר מעץ ההופמן הקנוני:

עץ השלד המתקבל מעץ ההופמן הקנוני	עץ השלד המתקבל מעץ ההופמן שאינו קנוני
קיבלנו 9 צמתים	קיבלנו 7 צמתים

מ.ש.ל.