

**תרגיל 9 במבני נתונים ואלגוריתמים**

**88-280 סמסטר א' תשע"ח**

1. דחסו את הטקסט הבא בעזרת אלגוריתם למפל זיו:

0110001110100110101010100

ב. וודאו שאתם יודעים כיצד לפתוח את הדחיסה.

**פתרון:** נקבל סך הכול שהמילון הוא:

0	0
1	1
01	2
11	3
10	4
00	5
001	6
111	7
101	8
100	9
011	10
1010	11
010	12
0101	13

ונקבל שהדחיסה היא:

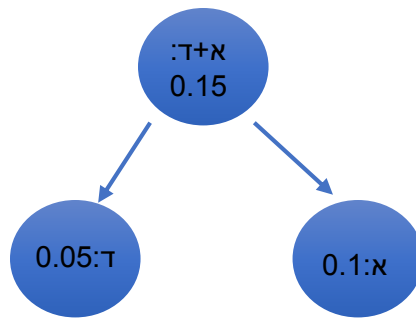
0	1	1	0	00	11	10	10	01	101	01	010	100
0	1	1	0	5	3	4	4	2	8	2	12	9

2. נתונה התפלגות מילים:

הסתברות	Word
0.1	א
0.2	ב
0.15	ג
0.05	ד
0.25	ה
0.13	ו
0.12	ז

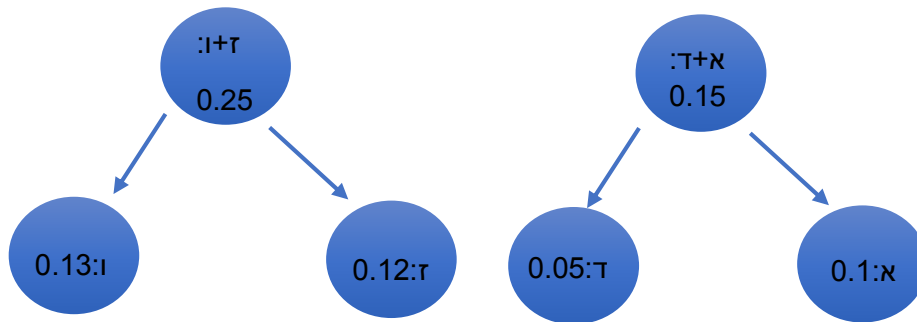
א. מצאו קוד הופמן עבור השכיחויות הנ"ל.

**פתרון:** נעבוד לפי שלבי האלגוריתם. נחבר קודם את א' וד' ונקבל מילה חדשה עם הסתברות 0.15.



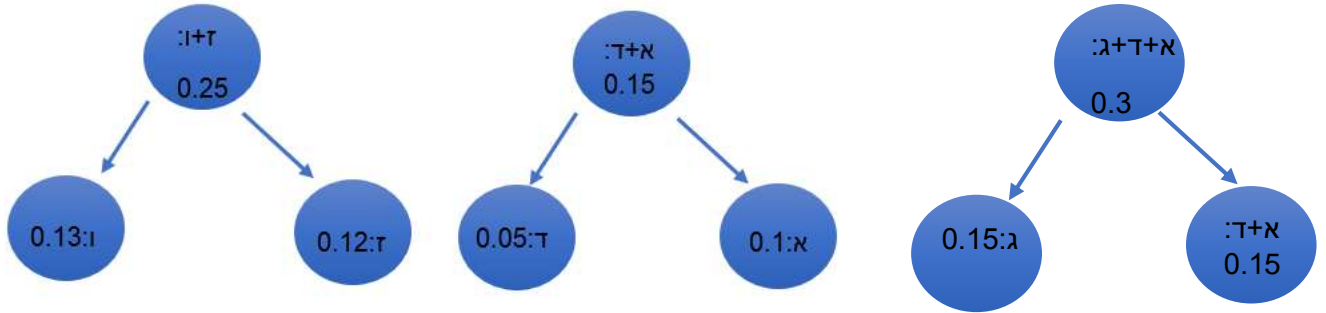
הסתברות	Word
0.2	ב
0.15	ג
<b>0.15</b>	<b>א+ד</b>
0.25	ה
0.13	ו
0.12	ז

בשלב השני נחבר את ז ו-ו' ונקבל מילה חדשה עם הסתברות 0.25



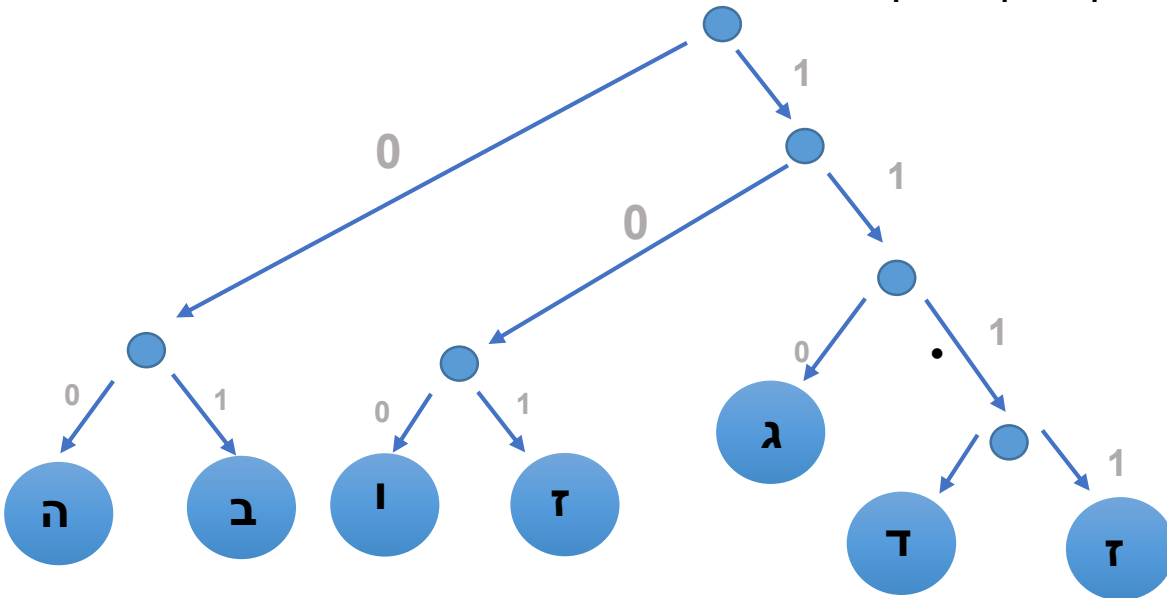
הסתברות	Word
0.2	ב
0.15	ג
0.15	א+ד
0.25	ה
<b>0.25</b>	<b>ז+ו</b>

בשלב השלישי נחבר את א+ד עם ג ונקבל א+ד+ג עם הסתברות 0.3.



הסתברות	Word
0.2	ב
<b>0.3</b>	<b>א+ד+ג</b>
0.25	ה
0.13	ו
0.12	ז

ונמשיך ככה ונקבל בסוף את העץ:



וקיבלנו:

א	0.1	1111
ב	0.2	01
ג	0.15	110
ד	0.05	1110
ה	0.25	00
ו	0.13	100
ז	0.12	101

ב. חשבו את כמות הביטים המינימלית הדרושה לייצוג אות בשפה.  
**פתרון:** זה בעצם לחשב את האנטרופיה.

H=-

$$(0.1 \cdot \log(0.1) + 0.2 \cdot \log(0.2) + 0.15 \cdot \log(0.15) + 0.05 \cdot \log(0.05) + 0.25 \cdot \log(0.25) + 0.13 \cdot \log(0.13) + 0.12 \cdot \log(0.12)) = 2.672$$

ג. חשבו את אורך המילה הממוצע לאחר הדחיסה.

**פתרון:**  $L = 0.1 \cdot 4 + 0.2 \cdot 2 + 0.15 \cdot 3 + 0.05 \cdot 4 + 0.25 \cdot 2 + 0.13 \cdot 3 + 0.12 \cdot 3 = 2.52$

3. אלגוריתם BM:

א. מצאו את טבלת BCT: ABABOBOBA.

**פתרון:**

B	O	A	*
1	2	0	9

ב. מצאו את ה GST עבור המילה מסעיף א'.

**פתרון:**

A	B	A	B	O	B	O	B	A
8	8	8	8	8	8	6	8	1

ג. תארו את שלבי החיפוש של התבנית BANANA בתוך הטקסט ANANASANANASANAS בעזרת BM. (רמז: חשבו את ה BCT).

פתרון: נחשב את ה-BTC:

A	N	B	*
0	1	5	6

נתחיל להשוות מול המילה ANANASANANAS.

A	N	A	N	A	S	A	N	A	N	A	S	A	N	A	N	A	S
B	A	N	A	N	A												
						B	A	N	A	N	A						
												B	A	N	A	N	A

הערה: נזכר שבאלגוריתם בודקים את המילה מול התבנית מהסוף. (מהאות האחרונה של המילה).

4. חשבו את טבלת ההסטות עבור המילה NANONANINAN של האלגוריתם KMP

פתרון: מחשבים כפי שנלמד בכיתה ומקבלים:

N	A	N	O	N	A	N	I	N	A	N	
-1	0	0	1	0	1	2	3	0	1	2	3

5. א. הכנס לעץ סיפא את המילה הבאה וצייר את העץ סיפא (בעת הבנייה יש צורך לצייר את העץ רק כאשר מתווסף פיצול): 011001100110. אין צורך לצייר SL. (מועד א' תשע"ז).

ב. הכנס לעץ סיפא את המילה הבאה וצייר את העץ סיפא: BANANA.

ג. הכנס לעץ סיפא את המילה הבאה וצייר את העץ סיפא: BANANABANANA.

פתרון: בדפי עזר.

6. נתון טקסט T מאורך n וסט של תבניות P1, P2...Pk מעל אלפבית סופי Σ, כל תבנית היא מאורך m. מצאו את כל האינדקסים ב-T כך ש:

$$T[i, \dots, i+m-1] = P_i[0, \dots, m-1]$$

עבור כלשהו  $P_i$  כלשהו בסט תבניות. עליכם לכתוב אלגוריתם ביעילות  $O(\max(mk, n))$ .

פתרון (בקצרה): נוכל להחזיק טבלת ערבול מגודל  $O(k)$ . נרוץ על הטקסט, נפעיל את פונקציית הערבול על כל תת טקסט מתאים ובדוק האם הוא מתאים לתבנית כלשהי בטבלת ערבול.

הוכחת נכונות והוכחת זמן ריצה – ש.ב.

8. מצאו בזמן לינארי ובעלות זיכרון לינארית את האלמנט הכי שכיח במערך.

פתרון (בקצרה): נשתמש בטבלת ערבול. נרוץ בלולאה על האיברים. עבור כל איבר נשאל – האם הוא מופיע בטבלת ערבול? אם כן, עדכן את ה-value שלו בטבלה ל-value+1 ואם לא, עדכן value=1.

כשנסיים לרוץ האיברים, נרוץ על כל ה-value בטבלת ערבול ונראה למי יש את ה-key המקסימלי (שבעצם זה מספר הפעמים שראינו את האיבר).

הערה: ניתן גם לוותר על הלולאה האחרונה בכך שיהיו לנו שני משתנים של הערך המקסימלי והאיבר הכי שכיח ונוכל לעדכן אותו כל הזמן במהלך הלולאה.

הוכחת נכונות והוכחת זמן ריצה – ש.ב.

