

16/11

נושא 4 - קוד האטמן (Huffman)

א) הדברה

קוד האטמן זהו קוד סבונה את כל הקידוד למטה עמדה. זה הוא מתחיל מסדר ההסתברות של התווים בסדר יורד. לא בהכרח ת"ב להיות הסתברות גלגל יכל להיות גם שכיחות. באופן כללי נקרא ערכים אלו "משקלים". כל איטרציה הוא מאחד שני קבוצות תווים לקבוצה אחת, כך שלב אחר ד-ח איטרציות מתקבל כל פנאלי שלם, הנקרא "קוד האטמן", נוצר את הקידוד.

Huffman(Σ)

$n = |\Sigma|$

$Q = \Sigma$

for $i=1$ to $n-1$ do

 create node z

$z.left = x = Q.extract_min()$

$z.right = y = Q.extract_min()$

$w(z) = w(x) + w(y)$

$Q.insert(z)$

return $Q.extract_min()$

// Q הוא תור עדינות או נרצות
מינימום המסדר והקטן ערכים

// שלף את שני הקטנים המנוצעים
ומאחד אותם לקבוצה z

// מחבר את משקלי הנצבים ב- z

מחזיר את שורש התבצולי //

כיצד ע"י יצירת:

תור	איברי
$\frac{1}{Q}$	1
$\frac{2}{Q}$	2
\vdots	\vdots
$\frac{n-1}{Q}$	$n-1$

ע"י קבוצה 2
שטחית ב- Q
ע"י Q היה שלב

סיבוכיות: יצירת העימוד Q היא בסבוכיות $O(n)$ וכך פונקטור הכנסה והוצאה בסבוכיות $O(n \log n)$. כיוון שהקוד בעצמו ד-ח נמצא, סיבוכיות העיודתם קוד האטמן היא $O(n \log n)$.

הצגת שיטת: בצדדים עשאו את קבוצת התווים בעזרה ניציה שני תווים הנעים A ו- B . תור A יחזיק

את כל המשקלים ההתחלתיים של התווים מאו"נים. הקטן הראש התור והגדול בסבוב. תור B יוצר עסמואים של המשקלים שאט אחדים כל איטרציה. כל איטרציה כדי עשלו את שני הקטנים עם המשקל האינצלי מסתק עשקוד 3 אנשח"ית: שני הגדלים הראשונים ב- A שני הגדלים הראשונים ב- B , הגדול הראשון ב- A ו- B . כל איטרציה נשא אק גלגל, אק ק נבחר (גלגל). אחת, נשא אק גלגל. אק ק נבחר (גלגל). אחת נבחר (גלגל). שני הצדדים נבחר יתאחדו לקבוצה z שתיכנס עסול תור B .

Ⓢ באיזה והתוים כבר מאו"נים בסדר עולה ע"י אשדום אעזריתם זה יתן עט סיבוכיות ע"י $O(n)$. אולם אק נצטרק ע"י את התור A בהעלה נקבע אותה סיבוכיות $O(n \log n)$.

ב) הוכחת אינצלי

נוכח שהקוד המוחזר מאעזריתם האטמן אק עקיים: $\sum_{i=1}^n w_i$ אינצלי, אק קודם נוכח 3 טענות עזר.

א) עמדה 1: קוד אינצלי תא"צ מיוצר ע"י צ" ע"י פנאלי מלא, שבו עכל צואת נניצית יש שני ערכים. הוכחה: נניח כשעדה שק"ם קוד אינצלי C שבו צואת נניצית B עם יעצ אחד. אזי נכל עיצור קוד חדש C' שוא כזה ע- C אלא שקעמנו את האב של B עם הקן של C בעקומו. נקבע ט- C עם קוד כמו C אק עביר אעדת קוד אחת עם ב"ל אחד נחות, בעתירה עאינצלי של C .



(44)

(2) למה 2: בע"פ בינארי המייצג קיבץ אינטרמדיאטי שני התווים בע"פ האינטרמדיאטי הנמוכים

- ביותר w ו- w_{n-1} , נמצאים ברמה הנמוכה ביותר בע"פ.

הוכחה: נניח על דרך השלילה שיש לנו ברמה הנמוכה ביותר בע"פ, אלא יש w ,

כך ש- $w > w_x$, והוא ברמה הנמוכה ביותר בע"פ. נמצא כי $w > w_x$. נבנה ע"פ

חדש z שהוא: בדיוק כמו הע"פ האינטרמדיאטי c אלא שהחלטנו בין w ו- w_x . מבחינת

יעילות השני הוא רק ב- w ו- w_x שני דגמים אלו. בע"פ c נחשב $w_x + w - w_x = w$

וגילו ב- c נחשב $w_x + w - w_x = w$. נבדוק איזה חישוב קטן יותר:

$$w_x + w - w_x = w > w_x + w - w_x = w \Rightarrow w_x + w - w_x = w > w_x + w - w_x = w$$

נמצא כי c יותר מינימלי מ- c , בסתירה לאינטרמדיאטי של c .

(3) למה 3: בע"פ אינטרמדיאטי ניתן להניח כי שני התווים בע"פ האינטרמדיאטי הנמוכים

ביותר w ו- w_{n-1} הם אחים.

הוכחה: דפי למה 2 הוא ו- w_{n-1} בוצא ברמה הכי נמוכה בע"פ, אלמס ע"פ בהכרח

אחים. נניח בשלילה כי הם לא אחים, אזי דפי למה 1 בוצא יש דשניהם

אחים אחרי w ו- w_x , שזם הם ברמה הכי נמוכה בע"פ. ניתן להחליף

בין w ו- w_x או בין w ו- w_{n-1} וכן דקבע ש- w ו- w_{n-1} בן יהיו

אחים בע"פ עשית את יעילות הקיבץ, שהרי $w_x = w - w_{n-1} = w$.

נחזיר להוכחה: נוכח באינדוקציה שעבור n תווים עם משקלים w_1, w_2, \dots, w_n , חישוב

היעילות $w_1 + w_2 + \dots + w_n$ דפי הע"פ שנבנה ע"פ יציא אלקטריהם הפאן שיהיה חישוב היעילות של

ע"פ אינטרמדיאטי עצור אותם תווים עם אותם משקלים

בס"ס - עבור $n=2$ הטענה נכונה כי ק"ס רק ע"פ אינטרמדיאטי אחד שזם אלקטריהם הפאן יתן.

הנחה - נניח שהטענה נכונה עבור $n-1$ תווים.

383 - נוכח עבור n תווים. יהי d ע"פ אינטרמדיאטי עבור n תווים. דפי למה 2 ו-3

שני התווים עם המשקל הכי נמוך w ו- w_{n-1} נמצאים ברמה הנמוכה והם אחים. נבדוק

ע"פ T_2 הבהה T_1 רק אלא w ו- w_{n-1} , כך ש- T_2 יש $n-1$ תווים, כי

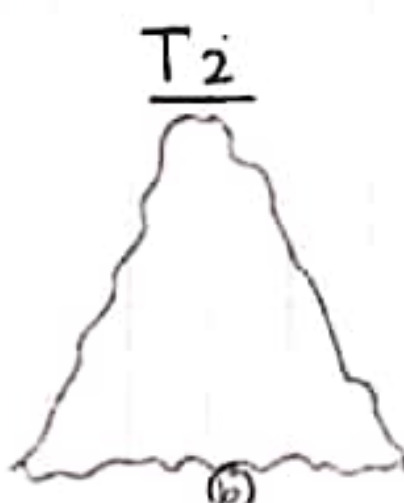
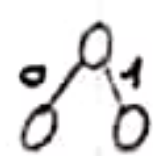
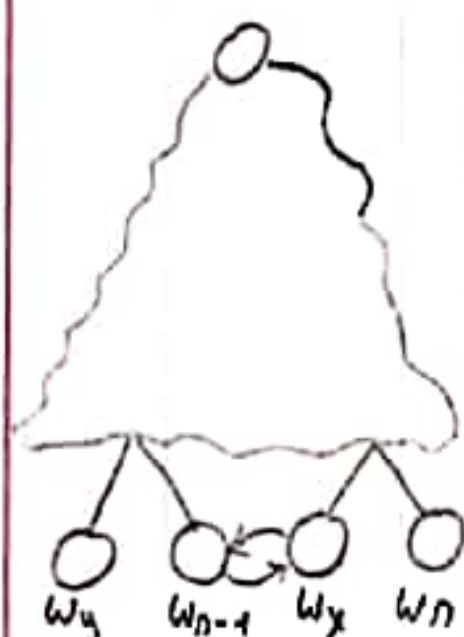
הורדנו 2 אק היססנו את האבא שלהם c . בנוסף את ק"ס: $w_b = w + w_{n-1}$. $w_b = w$ ו- w_{n-1} .

אם נוכח כי T_2 זק אינטרמדיאטי, אזי דפי הנחת האינדוקציה ניתן לבנות ע"פ אינטרמדיאטי אלקטריהם

הפאן T_1 שהיעילות שלו שווה ל- T_2 . דכן אק היינו מעבירים את האלקטריהם הפאן ע"פ התווים

היינו מבססים את T_1 עם w ו- w_{n-1} ברמה הכי נמוכה, דפי למה 2. ואז היה שקול ל- T_1 .

ג. תדירות יותר גבוהה.



46

33

 \Rightarrow

next code :

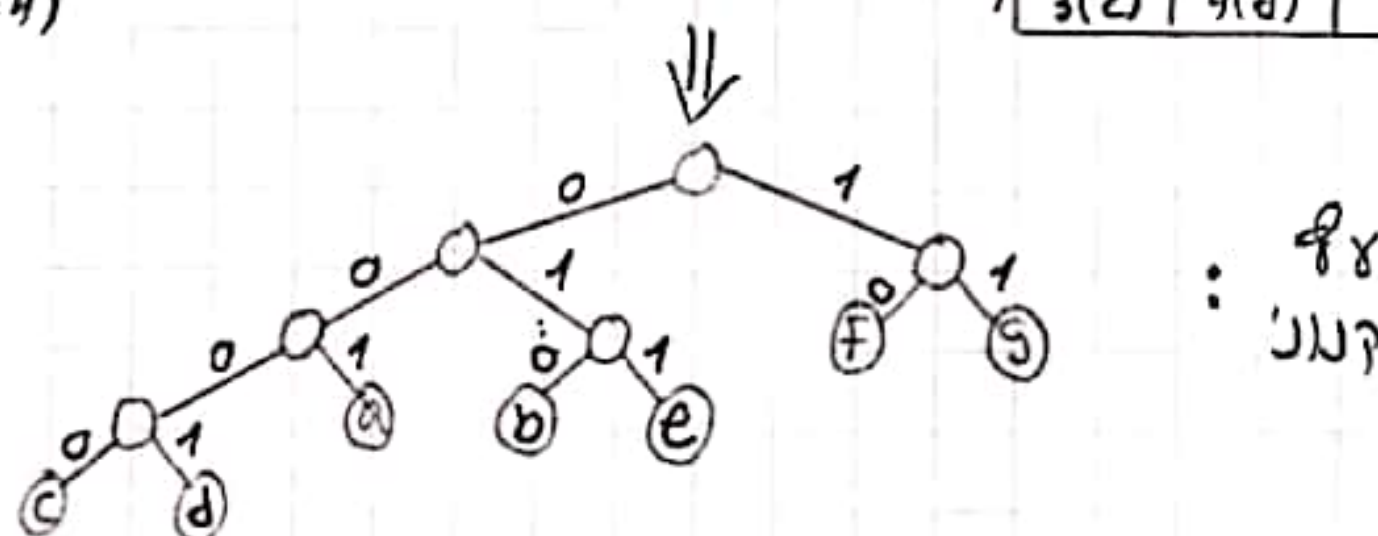
	1	2	3	4
2		2	3	4
		3	4	2
			4	2

(max length = 4)

1(a)	2(b)	3(c)	4(d)	5(e)	6(f)	7(g)
001	010	0000	0001	011	10	11

Symbo]:

	0	1	2
1			
2	6(f)	7(g)	
3	1(a)	2(b)	5(e)
4	3(c)	4(d)	



כ"ז ע"ב, צ"ל צ"ד ע"ב

ביגארי טראגטן בסוף סוף קינדס? אשתנה מאקרה מאקרה ותרלו ביצד האם, אטער העלדקיס ודירת הסף הקטן.

2

```

3 return symbol[L, v-firstcode[L]]

```

בואו לה: הריף את האסא-יתם עם "010110001" וזאת שהיטענו הוא .bgd

D-ary Huffman Code (3

HUFFMAN '30 '03 3'3'3' 8'6'02 02 7/08 מ'ה 15.

48

Sibling Property

שם מזה"ס את "תולדות האחים" עכ"ל:

נאמר
 הלאר אלאהם
 האנן דנא
 דנא דנא
 דנא דנא
 skeleton

(4) משקל של צומת פנימית שווה לסכום המשקל של שני ילדיו. $weight(w) = weight(w.left) + weight(w.right)$.

(2) אַז נאָר פֿאַר דעם האַנטלעךס פּאַסטער, מיט אונזער "נאָר", אַז ער פֿאַר אַ גאָנצן ייִדישן שולעךס באַגריפֿ.

$2j$ - $j-1$ ק אחרים. בג'סוח אחר: כל האחים בעד מאיס ער'ס באנגקס'ס עירק'ס.

משפט: אם M תכונת האחים"אם M היא תכונה של M ו- M היא תכונה של M אז M היא תכונה של M .

אדם ור' תם האמן צ'נא'

זהו אגרותם שכל נאם שקורא לו מדען את הפחיתות שלו ויודע הם צדד בזה שהולך 38 השלם

הסדר האוטומי זעיר
האשקלים בעל
י"ה כזה אחר כזה
אשקלים אשקלים
אשקלים.

הניכתי ימאוי עס תכינת האחים*, כן שבאלי האמלזותיך תתקנא עס האסאנן. מה עושים באשר מתקנים

ביתו עראשונה? עסק כן בעל 383 בעל תא'3 יהיה צוית בעל NYD מאשקו 0 וא"ו נטר

Not Yet Transmitted

את התו החדש ונעזבן כמו שנראה בהמשך אצל מלחינים. בתחילת האמרי"ם, העץ ואינך עץ חיים

נתגלה את כוונות האסטרות כי 383, שהוא קריאה של χ_2 :

$q_r = \text{I}e q F(x_{\pm})$ // $q_r = N Y T f_{\text{so}} z_s x_{\pm}$ p.t. x_{\pm} הנח $\rightarrow k$ צי"ח ה סני ה הזל q_r

$$15(q = NVT) \{$$

create left and right NYT children's to q

$$2 \quad q = q_{\text{left}} \pm$$

if (q is sibling to NYT node) Σ

Interchange a with highest numbered leaf of the same weight // wt 18 is 1887 as 18

q. weight ++

$$r = r_{\text{parent}}$$

```
while (ar ≠ root) {
```

Interchange u with highest numbered node of the same weight //

q. Weight ++

ע.מ.ק. ק"ל ז' חז"ל ואצ"כ"ס אמת חז"ל //

$$3 \quad q_r = q_{r, \text{parent}}$$

כחן 3871 המורה ++ weighty

[illegible]

ז' אב ה'תש"ז

אשר והא' נק'ם עלו בס' ד' א'

האינטואיציה של האשתיאס
הוא הפיזיקלי.

kalite ve dayanıklılığına göre

לפי ציור האפארייתם האמיתית
שחבונת האחים נשצרת.

2015/03

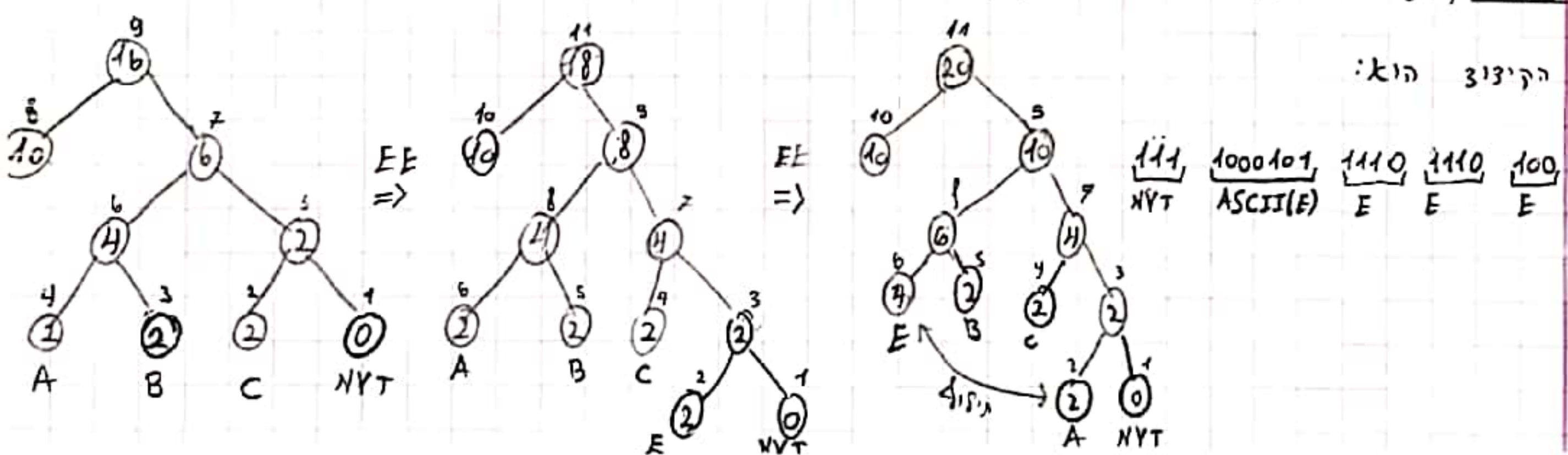
הקצו שטר באמארי'תם זה הוא שכל כסף שקורא'ץ מן הקובץ מקוצצו פ'ס' האס'ס'ס' 38 8-ז'א

אברהם אבינו ז"ל. הנה נהנה לך ביום הזה. וְהָיָה לְךָ לְחֵן וּלְחַן בְּיָמֶיךָ.

נקיטת גישה ה- Ascend על אף כבי' שהמסמך י'ז מותיר ג'יה תו חזק נכנס סוף.

בואמה: קיצי 3 אר הטקס הא T=FFFF כאשר הטקט שנסנה 38 כה היא הטקט אטלס:

הקיצוץ הוא:



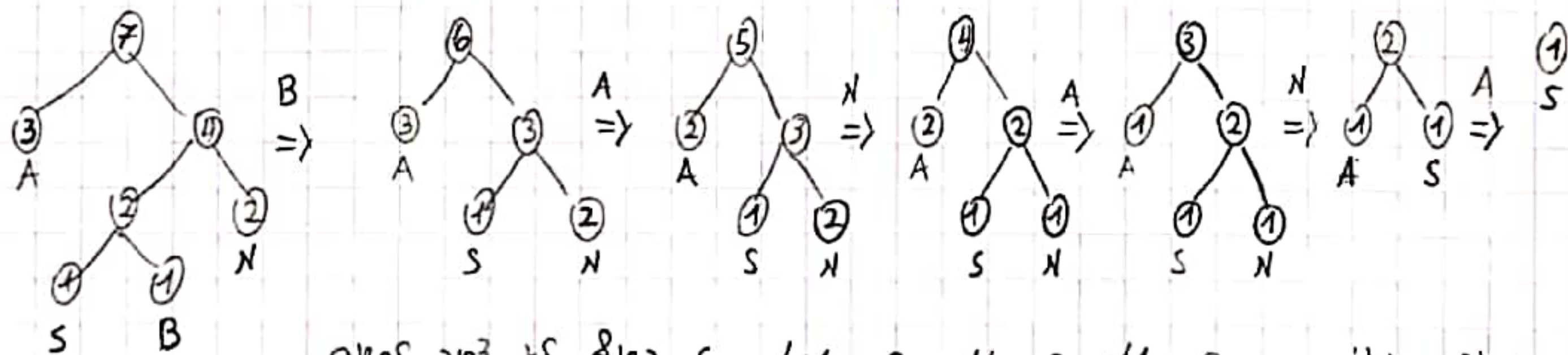
(1)

Forward Dynamic Huffman

זהו אלגוריתם שפיתח סטודנט של דנה שצובצ בקרינה הטובה על אלגוריתם שראינו בסעיף קודם. הוא מתחיל מעץ המאן סופי עם הקוסף, פשוט נסרקו הטבלאות והוסעם אלגוריתם המאן, ובסוף נוצר עץ הקוסף ומקוצצו תוך כפי האפשר. ברגע נדק שנתקם ברגע 5 מוריצ מהאשקף שבו 33: שמתאכנס וואז נמתק מהעץ, כך אנו מקדחים את העץ ואת הקיצוצו תוך מזהר עם הקוסף, נאמן יוצא קוצו קצר יותר מהאמאן נגיל.

בקיצוצו הוא עדי
מפסל ופשוט עשה
המייצג את התו בעץ
הנכתי של שם, כמו
בסוף קודם.

T = BANANAS



הקוצו הוא: $\frac{101}{B}$ $\frac{11}{N}$ $\frac{11}{A}$ $\frac{11}{N}$ $\frac{11}{A}$ $\frac{11}{N}$ $\frac{11}{A}$

אכן זהו התו היחיד שנשארה עץ טבלאות. צריך עטפות ה-Header טבלאות של כל האב.

הקיצוצו שנוצר באלגוריתם זה הוא טוב יותר מהאמאן הישן בעטפות 1-M ביטוי. נאמר מ זה אצל האב.

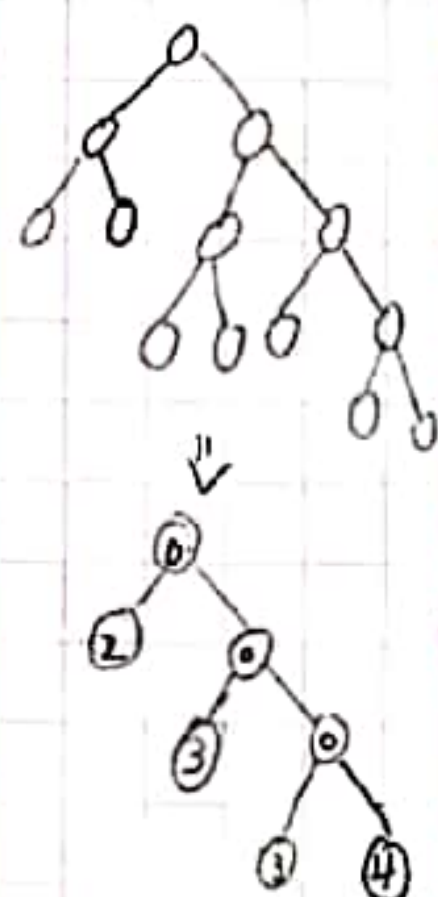
במצב 7 שקדים 31-32 יש השוואה בין אלגוריתם זה, המאן דנאי והמאן הישן.

Skeleton Tree

בסעיף 2 פסיד זה הדגנו אלגוריתם שיהיה על האמאן קטני. בסעיף זה ובסעיף הבא נלמד כיצד עטור את ידנו על זה, כך שנוכל לקצר אותו ולעטור אותו יותר מהר. בסעיף זה נלמד על "עץ טבלאות" ונראה כיצד יחסית ממוקמת על עץ האמאן קטני. יש לנו טבלאות ביטיות על טבלאות:

האפיקים: עץ האמאן קטני ואפיקים אלו בק טיפיה צטק יאניה באדוק טבלאות. עטם כך יש עתפיק כל ביל

בגם: קיצוצו 1-0 0-1 1-0 0-1 1-0 0-1 1-0. עץ זה שקול עץ האמאן קטני.



(2) נחשב: בעץ שהתקבל בני-עצים טבלאות שאתם נחתק. נגיש לב שאיך הקיצוצו בגם התוים שהם בתת-עץ

שם הם האלו אורק. ניתן משהם עכס הדמיתם בעץ שהתקבל, כך שכל צומת טבלאות מוקדח 0, וכל

ענף מוקדח יהיה אורק. קיצוצו של התוים שמחפדים אלו בקץ האמאן קטני המקורי. נאמן מוקדח צומת 1-M (Me)

בגם הביטוח יוכל עתפיות באמצעות עץ הטבלאות כך שכל בעס שגיל עשה בקץ נדע בקדק כמה עיצו ביל

עצמו לקדח, כך שנוכל לקדח אותם בבית אחת ולעטור את הקוצו באמצעות האלגוריתם שנלמד בהמשך. קריאה

של אסתי ביטוי בבית אחת ולא ביל אחת ביל חוסכת האין בעטיות ציטוק. העעקב קודא בגם נדק 8 ביטוי

(אנצט) ובסעיף עליהם בעטיות 1-M ו-0. עקב כל ביל בנכרצ. קריאה של אסתי ביטוי חוסכת בעטיות

אלו. יתרון טסל, שאכל הוף הוא אשמעותית יותר קטן כך שאנו שפחים בחיור מידע ה-Header עמנו

(20)

הגדרות: עטני שנספיר את האלגוריתם למענה נכח מספר הגדירות מהן האלגוריתם משתמש.

כזכור, בעץ האכסון קנוי ניתן עמק את כל גתוליים עמקוקים עטני אורק הקידוצ שלהם. עקוק ו

יכע את כל התוליים שאורק הקידוצ שלהם ו. בטסס, בתוק כל עקוק כל מילות הקידוצ

הם. מספיים בינאריים עקוקים.

• ון - מספר מילות הקידוצ שאיכס הוא עקוק ו. עמקים אחיות, אודע העקוק ו.

נסמן ב- m את אורק הקידוצ העממלי: $\{0 < m; i \in \mathbb{Z} \mid m = m\}$.

• $base(i)$ - עקוק העשירי של המספר הבינארי המייצג את הקידוצ של התו הראשון בעקוק ו.

נשים עכ כי עקוצ $base(m) = 0$, ובטסס $base(i) = 2(base(i-1) - 1)$, שכן אפ אוספים עייצוי העשירי

של מילות הקידוצ הראשונה מהעקוק הקידוצ את אודע אורא עקוק, ואז מכעעים ב-1 כדי עיוס

0 בטסס, מניעים עקידוצ התו הראשון בעקוק הבא.

• $B_s(k)$ - זהו הייצוג הבינארי של המספר k המשתמש ב- s ביטים. במידה והייצוג קטן מ- s אוספים אפס

בהתחלה. $B_4(3) = 0011$, נשים עכ כי $B_i(base(i))$ יתן עטני את הקידוצ של התו ה- i בעקוק ו.

• $seq(i)$ - זהו האינדקס של מילת הקידוצ הראשונה בעקוק ו ביחס עטאר מילת הקידוצ נשים עכ כי $seq(m) = 0$

ו- 1 , $seq(i) = seq(i-1) - 1$, שכן האינדקס של העקוק הקידוצ ואודע העקוק ניתן את האינדקס הראשון בעקוק הבא.

• $I(w)$ - זהו עקוק העשירי של אחיות בעליות w . מתעלם מאפסים בהתחלה. בעולה זו היא בעולה הוכי

ע- $B_s(k)$. נשים עכ כי $I(w) - base(i)$ כגשר w באורק ו יתן עטני את האינדקס של מילת הקידוצ w ב

עקוק ו. אפ נוסף עקוק זה את $seq(i)$ נקבע את האינדקס של מילת הקידוצ w ביחס עטאר מילת הקידוצ

• $diff(i)$ - מונדר עהיות $base(i) - seq(i)$. מתקיים: $diff(i) = I(w) - base(i) + seq(i)$, עכ, אורק זה עכ

כן מייצג את אינדקס מילת הקידוצ w ביחס עטאר מילת הקידוצ.

Decoding Algorithm

אלגוריתם למענה

$tree_pointer = root$

$i = start = 1$

while $i < length_of_string$ // כע עוק עכ נאיר הקידוצ

if $string[i] = 0$

$tree_pointer = tree_pointer.left()$

else

$tree_pointer = tree_pointer.right()$

if $value(tree_pointer) > 0$

$code_word = string[start, start + value(tree_pointer) - 1]$ // מודא את מילת הקידוצ עממח

$output = table[I(code_word) - diff[value(tree_pointer)]]$ // מונמח את התו עכ אינדקס המילה

$start = start + value(tree_pointer)$

$i = start$

// רכנית עממנות המילה הבאה

$tree_pointer = root$

else

$i++$ // עמיר עכיל הבא

}

(21)

(n)

[illegible]

התת-88 ש'372 מהס. אשה צדתי'ס ז'סו 'היה ח. ההטן מפין שט המשתדלים ס'ס'ס'ס.

באילנים, אחרית, אהבה, האסאן, קנול, את, כל, הקצתים, הכנ"א"ם, שאירק, כל, אילות, הקיץ, שו"צאות, ח"ק

הן בהכרח שם ו שם היותר ויותר שם את כל היתר שם שיוצא מהם.

בנימין פורמאנ' י"תר - פ' השלח 3 פ' ויחי 4 את האשכולים הפא' :

lower(v) = upper(v) = val_u(v) יט ה י ו ז ח ט

• v ו- v' נמצאים באותו pk . $upper(v) = upper(v'.right)$, $lower(v) = lower(v'.left)$. מכאן נובע שהאנדרטות

ה'תש"ס שו"ת א"ק מ-י. פודקיס את האשכנזי גבור בן יאני כיוון שהחלף שם אשכנזי יאני.

$\text{lower}(w) = \text{lower}(w)$ $\text{upper}(w) \leq \text{lower}(w) + 1$ $\text{upper}(w) \leq \text{lower}(w) + 1$

נסמן באישהו δ $Flag(v) = 1$ φ γ היא צורת איות 3 נא שיתאנו ע

במסגרת 6 ת"ת

37-38 נ'מ 87-א

את כל השאלות מאמר

356 488 512 488

מנדט.

Decoding Algorithm 1

```
tree - Pointer = root
```

$$i = 5 \pm 9r \pm 1$$

```
while i < length-of-string // loop ends
```

if string[i] = 0

```
tree.Pointer = tree.Pointer.left()
```

else

```
tree_pointer = tree_pointer.right()
```

```
IF Value(tree_Ponter) > 0 {
```

```
len = Value(&free_pointer)
```

code word = string[start, start + len - 1] // תחילת המילה

```
if Flag(tree_pointer) == 1  AWP  2.I(code word) > base(len+1) }
```

```
codeWord = string[start, start + len]
```

10011

$output = table[1(codeWord) - diff][len]$ // זהו הקוד של פונקציית decode

$$i = \text{star} \pm = \text{star} \pm + 1 \text{ en}$$

3 free_pointer = root // הנקודה הראשונה

else

עמוד עמ"ט הקא 11 111

3

הערה - האדמיותם '33' האם עקומה ה או 1-1 מיל'ס, ע"י בק שמוקד את ארשת הקוב באורך ה

ממנה אותה עשיתי ואיסוף זה כסוף. גם הערך שהתקבל גבוה מה"דני הנשרתי גם האמת ק"ב

הראשונה בעמ'יק באורן וזו אזי דייק עקרא וזו עמ'יק, אחרת, קורא'יק ח עמ'יק.