אלגוריתמים קומבינטוריים
סיכומים של תרגילי כיתה מסמסטרים קודמים בנושא
קודי הופמן Huffman Codes

1. הגדרה: קוד רישא הוא קידוד של האותיות (של אלף-בית של הטקסט) המקוריות על ידי מילים (מעל אלף-בית של הקוד (למשל בינארי)) כך שאף מילת קוד הוא לא ההתחלה של מילת קוד שונה, לקוד כזה קוראים קוד רישא.

   קוד כזה הוא קל לפענוח - קוראים אותיות על שנגמר מילת קוד, ורושמים את האות שהוא מקודד. מטרת קוד כזו הוא להעביר את הטקסט המקורי באורך קצר ככל שאפשר. לכן רוצים לקודד אות שמופיעהבתדירות גבוהה במילת קוד קצרה.

2. האלגוריתם של הופמן נותן דרך לבנות קוד רישא בינארי אופטימלי (ביחס להתפלגות נתונה של שכיחויות האותיות). נסתכל על האלגוריתם כשיטה לבנות עץ בינארי מלא שמתאים לקוד רישא. בכל עלה יהיה אות מהקלט, ומילת הקוד תוגדר לפי המסלול מהשורש לעלה שלו - כאשר נקח קשת לבן שמאלי נרשום 0, וכאשר נקח קשת לבן ימני נרשום 1. (מילת הקוד מתחיל מהאות שמתאים לקשת שיוצאת מהשורש ומסתיימת באות שמתאים לקשת שנכנס לעלה.)

3. תיאור האלגוריתם: נרשום כל אות כעלה, ובכל עלה גם נרשום את ההסתברות (או השכיחות) של האות בקלט. נחזור על התהליך הבא עד שנקבל עץ בינארי מלא (שכל צומת פנימית יש לו שני בנים) המכיל את כל העלים: נבחור שני אברים בעלי ההסתברויות הנמוכות ביותר וניתן להם אבא משותף שהסתברותו הסכום של ההסתברויות שלהם. את הבנים נוציא מרשימת האברים לבחירה ובמקומם נשים את האבא המשותף החדש.

4. בהרצאה הוכח שאלגוריתם זו נותן קוד רישא בינארי אופטימלי (ביחס לפילוג הנתון). גם ראינו שכל עץ בינארי מתאים לקוד רישא, אך רק עץ בינארי מושלם יכול להתאים לקוד רישא אופטימלי (ביחס להתפלגות מסוימת).

5. תרגיל (ממבחן ישן):

   (א) מצא קוד רישא (בינארי) אופטימלי לסדרת השכיחויות:

   $$\frac{1}{24}, \frac{1}{12}, \frac{1}{12}, \frac{1}{8}, \frac{1}{8}, \frac{1}{6}, \frac{3}{8}.$$

   (ב) מצא קוד רישא (בינארי) עם 7 מילים שאורכיהם $2, 2, 3, 3, 3, 4, 4$. מצא סדרת שכיחויות $p_1, \ldots, p_7$ עבורן הקוד הזאת הוא אופטימלי.

   (ג) האם קיים קוד רישא (בינארי) עם 7 מילים שאורכיהם $2, 2, 2, 3, 3, 5, 6$? אם כן, האם הוא קוד רישא אופטימלי עבור סדרת שכיחויות כלשהיא?

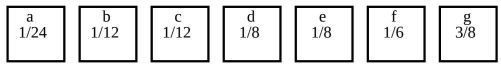
   (ד) האם קיים קוד רישא (בינארי) עם 7 מילים שאורכיהם $2, 2, 3, 3, 4, 4, 5$? אם כן, האם הוא קוד רישא אופטימלי עבור סדרת שכיחויות כלשהיא?

פתרון:

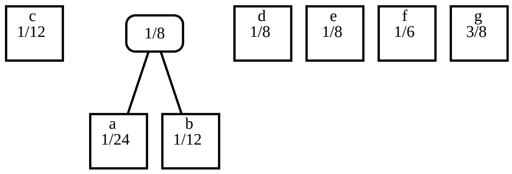
(א) נפעיל את האלגוריתם של הופמן כדי לבנות עץ בינארי מלא המתאים לקוד רישא אופטימלי. בציורים בדף הבא מופיעים השלבים השונים בתהליך. הקוד שהתקבל הוא:

$$a = 0110,\ b = 0111,\ c = 010,\ d = 100,\ e = 101,\ f = 00,\ g = 11$$

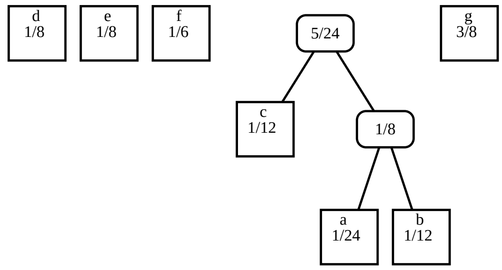
שלב 1:



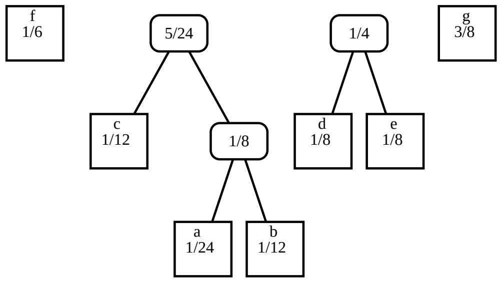
שלב 2:



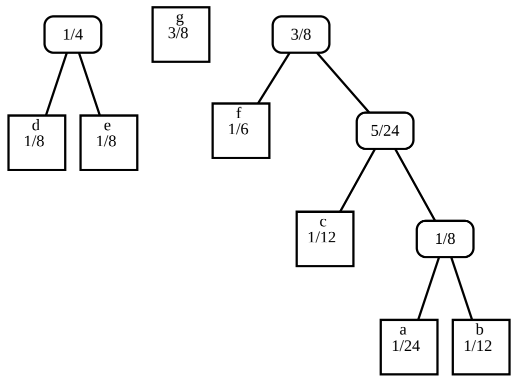
שלב 3:



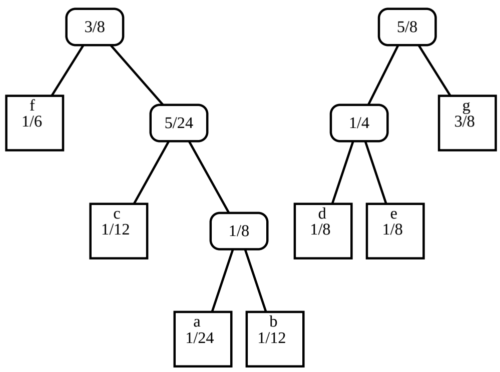
שלב 4:



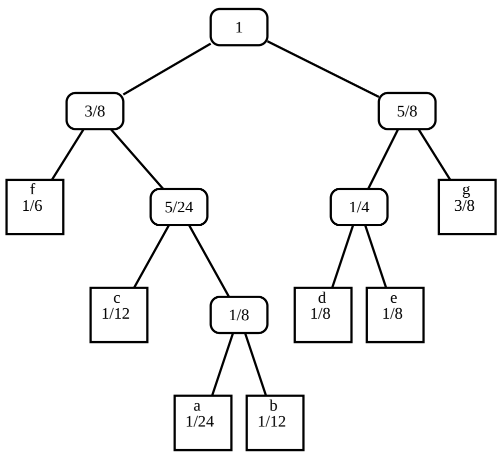
שלב 5:



שלב 6:

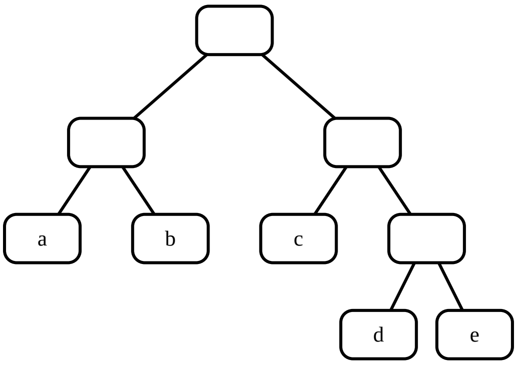


שלב 7:

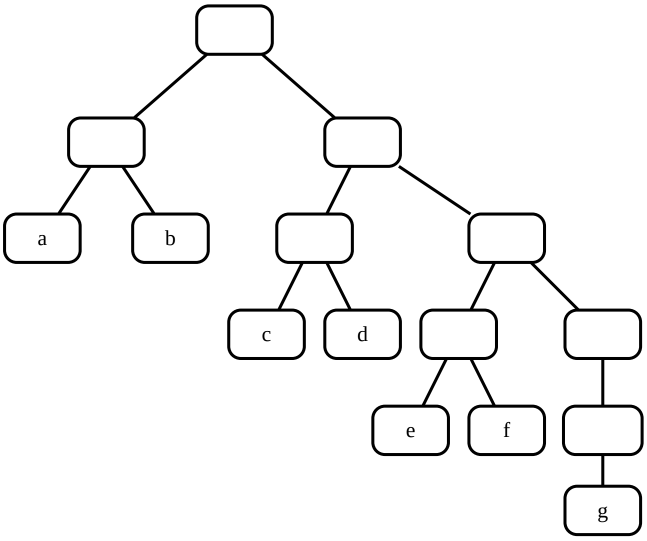


(ב) הקוד שהתקבל בסעיף א' הוא כנדרש. אם לא היה לנו קוד מתאים - היינו מנסים לבנות עץ בינארי מלא מתאים, ואחר כך מלביש עליו הסתברויות כך שהאלגוריתם של הופמן יבנה אותו. (ניתן לעשות כן מהעלים העמוקים למעלה... או ניתן להשתמש בחזקות מתאימות של $\frac{1}{2}$.)

(ג) לא קיים קוד רישא כזאת. כאשר ננסה לבנות עץ בינארי מלא מתאים, לאחר שנסמן שלושה עלים במרחק 2 מהשורש ושני עלים במרחק 3 מהשורש העץ הוא מלא ואין אפשרות להוסיף את העלים במרחקים 5 ו- 6. (ראה ציור.)



(ד) קיים קוד רישא כזאת אך הוא אינו מלא - ולכן לא יכול להיות אופטימלי. (ראה ציור.)



6. גישה אחרת לבעיה משתמש בתוצאה הבאה: יהיו $l_1, l_2, \ldots, l_n$ האורכים הרצויים של מילות הקוד בקוד רישא. אזי נתבונן בסכום:

$$S = \sum_{k=1}^{n} \left(\frac{1}{2}\right)^{l_k} = \sum_{k=1}^{n} 2^{-l_k}$$

- אם $S = 1$ קיימת התפלגות (מתאימה) וקיים קוד רישא באורכים הרצויים שאופטימלית להתפלגות הזאת.
- אם $S > 1$ לא קיימת אף קוד רישא עם מילות קוד באורכים הרצויים.
- אם $S < 1$ קיים קוד רישא באורכים הרצויים אך היא לא אופטימלית לאף התפלגות שבה כל ההסתברויות חיוביות. (לפעמים שימוש נכון בהסתברות 0 עוזרת, אך אם כן יש מילות קוד מיותרות!)