בע"ה

**דחיסה**

מספר הקורס: 89-745-01

סמסטר א מועד א

תאריך: 21.1.13 ; י שבט תשע"ג

מרצה: פרופ' שמואל קליין

משך הבחינה: שעתיים

*הנחיות*: א) אין להשתמש בכל חומר עזר, גם לא מחשבונים

ב) ענה/י על כל השאלות

ג) כל תשובה חייבת להיות מנומקת היטב

[1] (40 נקודות)

נתון הטקסט הבא, שאותיותיו לקוחות מתוך האלף-בית {a, b, c, d, e, f, g, h}

b d e c c g e a b d e c c c c f g a h d e f e a b d e c f d d e a b c c c c c f

התדירויות של האותיות הן:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| h | g | f | e | d | c | b | a |
| 1 | 2 | 4 | 7 | 6 | 12 | 4 | 4 |

1. בנה קוד הופמן 3-ary עבור אלף בית זה לפי תדירות הופעת האותיות בטקסט. השתמש בתוים A, B, C, כדי לציין את מילות הקוד על בסיס 3-ary , ואח"כ החלף כל אחד מהתוים במחרוזת בינארית באורך 1 או 2 כדי למזער את גודל הקובץ הדחוס.

ב) חשב את אחוז הדחיסה שהתקבל בסעיף א) יחסית להצפנת הטקסט בקוד שבו כל אותי מילות הקוד שווים.

ג) הצפן את הטקסט לפי שיטת Lempel-Ziv-Welch ובנה במקביל את הטבלה.

ד) מהו אחוז הדחיסה המתקבל (יחסית להצפנה ב fixed-length-code) אם המצביעים לטבלה מוצפנים.

1. ב fixed-length.

2. בצורה הדרגתית, כאשר מתחילים עם טבלה בגודל מינמלי ומכפילים את גודלה כאשר זה נחוץ?

[**2**] (40 נקודות)

נתונה קבוצה גדולה של מפות של סיביות, כל מפה בעלת אותו אורך N שהוא מספר גדול (למשל 50000) הבעיה שהן צורכות הרבה שטח, אך לרוב הן דלילות מאוד (מספר אחדים קטן מאוד לעומת מספר אפסים גדול). ראינו כמה שיטות לדחוס מפות כאלה, אך כל השיטות דוחסות כל מפה לחוד. ניתן אולי לשפר זאת אם דוחסים את הקבוצה כולה וניתן להשתמש במפה אחת כדי לדחוס אחרת.

(10) א) הראה/י שאם משתמשים בקידוד אריתמטי, מידת הדחיסה של מפה (compressibility) תלויה בדלילותה.

(10) ב) אם מפה A ומפה B דומות מאוד, אז המפה C=A xor B תהיה דלילה יותר מ-A וגם מ-B. איך נוכל לנצל זאת?

(20) ג) איך נפעיל את הרעיון ב-ב) על כל קבוצת המפות? הדרכה: צריך למזער את מספר הביטים הדלוקים במפות הרגילות ובאלה שנשמרים בעזרת xor.

[**3**] (20 נקודות)

קוד פיבונצ'י הוא סדרה אינסופית של מילות קוד

{11, 011, 0011, 1011, 00011, 10011, 01011, 000011, 100011, ...}

שבה מילות הקוד מאופיינות ע"י כך שכולן מסתיימות ב- 11 וכן שהמחרוזת 11 אינה מופיעה באף מקום אחר באף מילת קוד. בדומה לכך אפשר להגדיר קוד טריבונצ'י:

{111, 0111, 00111, 10111, 000111, 100111, 010111, 110111, 0000111, 1000111, 0100111, 1100111, 0010111, 1010111, 0110111, 00000111, 10000111, …}

שבה מילות הקוד מאופיינות ע"י כך שכולן מסתיימות ב- 111 וכן שהמחרוזת 111 אינה מופיעה באף מקום אחר באף מילת קוד. סדרת טריבונצ'י מוגדרת ע"י

T(0)=T(1)=1, T(2)=2. T(i)= T(i-1) + T(i-2) + T(i-3) for i>2.

1. מה הקשר בין סדרת טריבונצ'י לקוד טריבונצ'י?
2. דון/י ביתרונות וחסרונות של קוד טריבונצ'י לעומת קוד פיבונצ'י.

בהצלחה !!!