

מטלת מנחה (ממ"ן) 14

מטלה תכנותית

הקורס: 20407 – מבני נתונים ומבוא לאלגוריתמים

חומר הלימוד למטלה: חוברת הקורס וספר הלימוד

משקל המטלה: 5 נקודות

מספר השאלות: 1

מועד אחרון להגשה: 08.01.2021

סמסטר: 2021א

אופן הגשת המטלה:

שליחת המטלה תתאפשר רק באמצעות מערכת המטלות המקוונת באתר הבית של הקורס

הסבר מפורט ב"נוהל הגשת מטלות המנחה"

רצוי להתחיל לעבוד על המטלה לפחות שבועיים לפני מועד ההגשה

מותר לעבוד בזוגות, אך שימו לב להנחיות בהמשך

מבוא

בממ"ן זה עליכם לכתוב ולהריץ תכנית ב-Python, Java, או ב-C/C++¹, אשר מממשת מבנה נתונים מבוסס טבלאות-גיבוב כמפורט בהמשך.

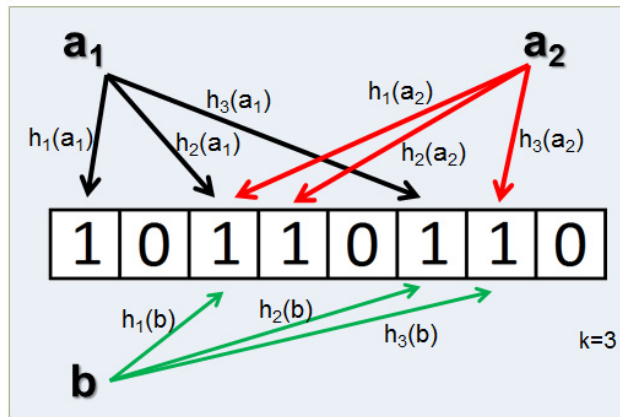
לפני שתמשיכו בקריאה, קראו תחילה את סעיף 2.6 בחוברת הקורס (הנחיות לכתיבת ממ"ן תכנותי).

ייתכן שיינתנו הנחיות נוספות לקראת מועד ההגשה – אנא בדקו באתר.

¹ למעוניינים להגיש בשפה אחרת, אנא בדקו תחילה עם המנחה שבודק/ת את התרגילים שלכם

כללי:

ברצוננו לבנות מערכת אשר תוכל לבדוק במהירות שייכות של איבר למבנה נתונים (חיפוש). כדי לבצע את הבדיקה במהירות (גם אם על חשבון דיוק) המערכת משתמשת בטבלה של **ביטים** בגודל m , וב- K פונקציות גיבוב (HASH) המחזירות כל אחת ערכים בין 0 ל- $(m-1)$. ניתן להניח שהפונקציות מגבבות באופן אחיד, כלומר הסיכוי של איבר להתגבב על ידי פונקציה מסוימת לתא מסוים הוא $1/m$. הטבלה T נבנית באופן הבא:



דוגמה: הכנסה של שלושה איברים ע"י שלוש פונקציות

באופן הבא:

אפס את m התאים של T

לכל איבר a_i השייך למבנה:

לכל פונקציה גיבוב h_j :

$\text{index} = h_j(a_i)$

$T(\text{index}) = 1$

בעת קבלת איבר a לבדיקת שייכות, המערכת מחשבת את K ערכי הגיבוב של איבר זה ומחזירה "כן" אם כל התאים המתאימים לערכים שהתקבלו דלוקים. כלומר:

if $h_1(a) == 1 \ \&\& \ h_2(a) == 1 \ \&\& \dots \ \&\& \ h_K(a) == 1$

return YES

return NO

א. מה הסיכוי שאיבר ששייך למבנה יוכרז כלא שייך?

ב. הראו שהסיכוי שאיבר שלא שייך למבנה יוכרז כשייך הוא: $(1 - (1 - \frac{1}{m})^{KN})^K$

הדרכה: הראו תחילה שהסיכוי ש- $h_i(a) = 0$ הוא $(\frac{m-1}{m})^{KN}$

ג. חשבו את השגיאה (כלומר הסיכוי שאיבר שלא שייך למבנה יוכרז כשייך) עבור:

$$K=13, m=32 \cdot 10^6, N=10^6$$

ד. ממשו את המבנה הזה לפי ערכים שמתקבלים מהמשתמש. את משפחת הפונקציות אתם יכולים לממש בעצמכם או להשתמש במימוש של *MurmurHash* המתאים לשפה שבה אתם כותבים.

צורת הקלט:

עליכם לקלוט את m -ו- K מן המשתמש (ניתן להניח ש- m ו- K לא גדולים מהערכים בסעיף ג') קליטת האיברים להכנסה למבנה תיעשה מתוך קובץ. יש לאפשר קבלת שני קבצים: קובץ המכיל איברים מספרים או מחרוזות (מופרדים בפסיקים) של איברים להכנסה למבנה. קובץ המכיל איברים מספרים או מחרוזות (מופרדים בפסיקים) של איברים לבדיקת שייכות למבנה.

מהלך התכנית:

עליכם לאפשר הכנסת איברים מתוך הקלט למבנה. עליכם להכניס אותם לפי האמור לעיל. אין צורך לשמור את האיברים עצמם (אם כי אפשר) אלא רק להכניסם למבנה המתואר לעיל. לאחר הכנסת האיברים, עליכם לבצע בדיקה של האיברים מקובץ הבדיקה ולהדפיס לכל איבר מקובץ אם הוא נמצא שייך או לא (לפי תוצאת הבדיקה). מערך הביטים יכול להיות ממומש כמערך של *integers* או של ביטים. ניתן להשתמש במימושים של מערך ביטים הקיימים עבור שפת התכנות שבחרתם.

צורת הפלט:

עליכם לאפשר הדפסת התוצאות למסך.

מסמך מלווה:

מעבר לפתרון סעיפים א' עד ג' עליכם לתאר במסמך מלווה את אופן פעולות שאר האלגוריתמים שמימשתם עם ניתוח זמנים ודוגמאות הרצה.

בונוס:

בונוס יינתן עבור ביצוע חישובי סטטיסטיקות ע"י הכנסת איברים לטבלאות בגדלים שונים ומספר פונק' גיבוב שונות וספירת אחוז השגיאות. לעובדים בזוגות, זוהי דרישת חובה (ולא בונוס). עם זאת, יינתן בונוס גם לזוגות ע"פ שיקול דעת הבודק למימוש יוצא דופן של ניתוח התוצאות.

כאמור, ייתכנו הנחיות נוספות לקראת מועד ההגשה – אנא עקבו באתר.