

מועד ג'

1 מתוך 1

מספר תעודת זהות:

מספר מחברת:

מספר קורס: 0368.2158

סמסטר ב', מועד ג', תשס"ז

28/12/2007

מבחן במבני נתונים

פרופ' חיים קפלן, אלעד ורבין, ליאור שפירא

משך המבחן שלוש וחצי שעות (לא תינתן הארכה)

הבחינה עם חומר סגור, ללא מחשבים/מחשבוניס. מותר להביא דף עזר אחד.

הוראות כלליות:

1. יש לרשום מספר ת.ז. ומספר מחברת בראש כל דף.
2. המחברות הן לטייטה בלבד ולא תיבדקנה.
3. מותר להשתמש באלגוריתמים שנלמדו בכיתה כקופסאות שחורות.
4. אין לכתוב אלגוריתמים בפסאודו-קוד אלא להסביר (באופן משכנע) את הרעיון הכללי.
5. כתבו תשובות קצרות ומדויקות. תשובות לא ממוקדנות וארוכות, מסובכות או מסורבלות יקבלו ניקוד חלקי גם אם הן נכונות. תשובה ללא נמוק במקום שנדרש נמוק לא תזכה בנקודות.
6. סיבוכיות זמני ריצה יש לכתוב כחסם הדוק ככל שתוכלו, בכת (ב) O.

בהצלחה!!

שאלה	ציון	ניקוד מקסי
1		20
2		16
3		16
4		16
5		18
6		14
סה"כ		100

שאלה 1 (20 נקודות)

ענו על השאלות הבאות בקצרה (משפט או שניים לכל היותר)

- א. יהיה X צומת המכיל מספר k בעץ חיפוש בינארי מלא, בו בכל צומת יש מספר. בהינתן כי $k+1$ נמצא בעץ ואין בעץ מספר y כך ש- $k < y < k+1$ היכן הצומת שמכיל את $k+1$?

- ב. באיזו סיבוכיות זמן ניתן לבנות ערמה בעלת צמתים?

- ג. קוטר עץ T הינו המסלול הארוך ביותר בין שני צמתים כלשהם. תנו חסם עליון על קוטר עץ אדום-שחור בעלת צמתים.

במקו: $O(\quad)$

- ד. נתון צומת v בעץ אדום-שחור $P1$ ו- $P2$ הם אורך המסלול הקצר ביותר מ- v לעלה ואורך המסלול הארוך ביותר מ- v לעלה בהתאמה. תנו חסם תחתון על היחס בין $P1$ ל- $P2$. נמקו את תשובתכם.

מועד ג'

3 מתוך 3

מספר תעודת זהות:

מספר מחברת:

ה. באחד ממבני הנתונים לבעיית ה-union find השתמשנו בשיטה הנקראת 'כיווץ מסלולים' (path compression), תארו את המבנה והשיטה בקצרה.

ו. בהינתן עץ T בערמת פיבונצ'י, כך שלשורש העץ יש דרגה k. תנו חסם תחתון על מס' הצמתים בעץ, וחסם עליון על k

מס צמתים = $\Omega(\quad)$
k = $O(\quad)$

במקו:

שאלה 2 (16 נקודות)חלק א'

נתון עץ אדום שחור T (עם הערכים בעלים ועם מצביע מכל צומת לאביו). מבצעים m פעולות הכנסה לעץ כאשר בכל פעולה נתון עלה שלפניו או אחריו יש להכניס את האלמנט החדש. הוכיחו כי עלות סדרת הפעולות הינה $O(m)$ (ראו נספח עם חמשת המקרים של $insert$ לעץ אדום שחור).

חלק ב'

עץ AVL הינו עץ חיפוש בינארי מאוזן המוגדר כך שעבור כל צומת T בעלת תת עץ ימני T_R ותת עץ שמאלי T_L , מתקיים כי: $|H(T_R) - H(T_L)| \leq 1$ כאשר H פ' גובה של תת עץ. הוכיחו כי גובהו של עץ AVL בעל n מפתחות הוא $O(\log n)$.

מועד ג'

5 מתוך 5

מספר תעודת זהות:

מספר מחברת:

שאלה 3 (16 נקודות)

נתונה המחרוזת S: BANANA

א. ציירו את ה-suffix tree של S.

ב. תארו אלגוריתם הרץ בזמן ליניארי המקבל כקלט מחרוזת S ומוצא את תת המחרוזת הקצרה ביותר החוזרת פעם אחת בלבד ב-S. לדוגמה במחרוזת agtggacatggg תת המחרוזת "ט" הינה התשובה הנכונה.

מועד ג'

6 מתוך 6

מספר תעודת זהות:

מספר מחברת:

ג. נתונות אותיות ותדירות הופעתן בטקסט. ציירו את עץ הופמן המתאים לקלט:

אות	א	ב	ג	ד	ה	ו	ז
תדירות	40	60	25	50	1	12	5

שאלה 4 (16 נקודות)

נתונה טבלת hash עם $m=11$ כניסות. כמו כן נתונות שתי פונקציות:

$$h_1(\text{key}) = \text{key} \bmod m$$

$$h_2(\text{key}) = \{\text{key} \bmod (m-1)\} + 1$$

א. הכניסו את המפתחות $\{11, 3, 7, 22, 18, 33\}$ לטבלה לפי הסדר (משמאל לימין) בכל אחת מהשיטות הבאות:

א.1. Linear probing כאשר התא ה- i שנבדוק למפתח k הוא $h(k, i) = (h_1(k) + i) \bmod m$

א.2. Double hashing כאשר h_1 פונקצית ה-hash ו- h_2 פונקצית המרווח. כלומר התא ה- i

שנבדוק למפתח k הוא $h(k, i) = (h_1(k) + i * h_2(k)) \bmod m$

	Linear Probing	Double Hashing
0		
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

מועד ג'

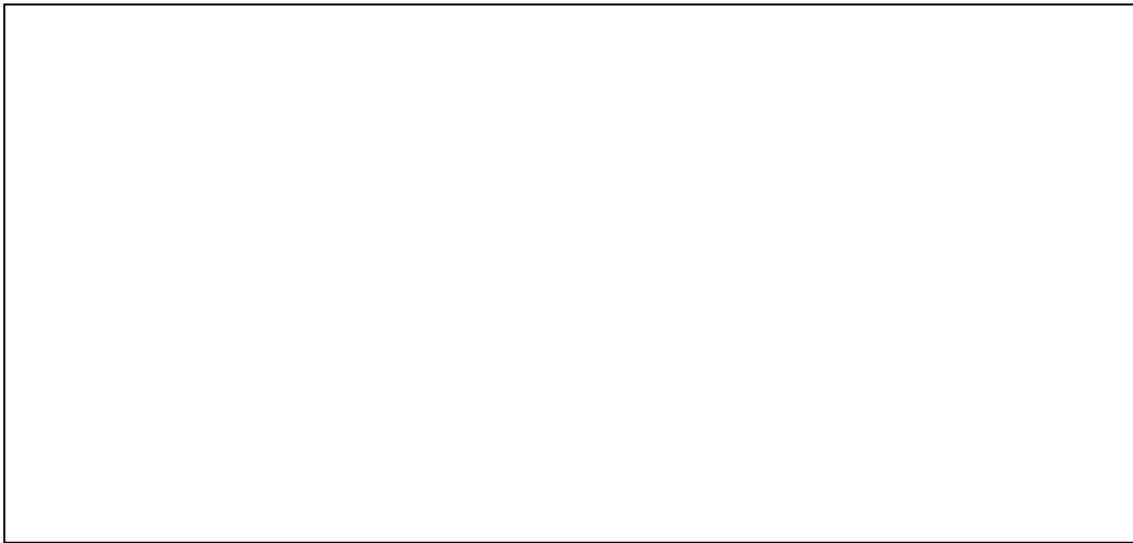
8 מתוך 8

מספר תעודת זהות:

מספר מחברת:

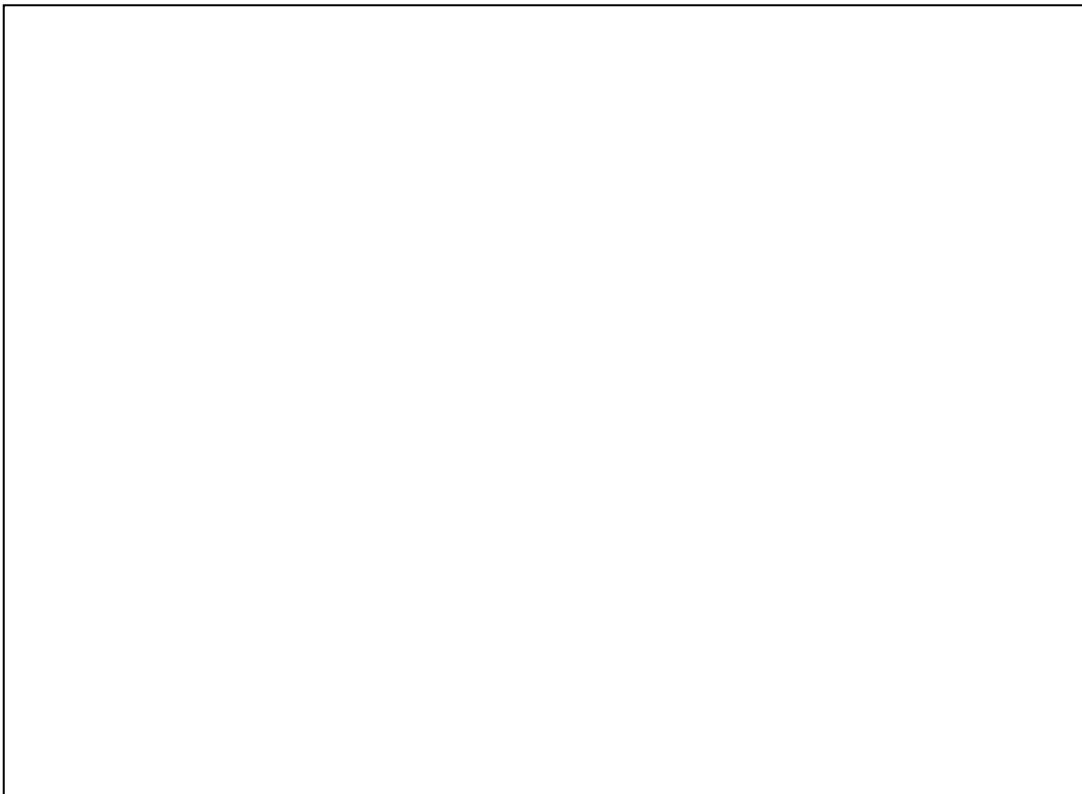
ב. מה יקרה אם בא 2. נשתמש ב- h_2 כפונקצית ה-hash וב- h_1 כפונקצית המרווח? כלומר

$$h(k,i)=(h_2(k)+i*h_1(k)) \bmod m$$



ג. הסבר מדוע כאשר $GCD(h_{step}(k),m)=1$ עבור כל k , התופעה בסעיף ב' אינה יכולה

להתרחש. הפונקציה $h_{step}()$ היא פונקצית המרווח.



שאלה 5 (18 נקודות)

א. בהינתן קבוצה של n מספרים שונים ומספר k (כאשר n מתחלק ב- k ללא שארית), תארו אלגוריתם יעיל ככל שתוכלו, שמחלק את הקבוצה ל- k קבוצות שוות בגודלן S_1, \dots, S_k כך שכל המספרים ב- S_i קטנים מכל המספרים ב- S_{i+1} לכל $1 \leq i \leq k-1$. מהו זמן הריצה של האלגוריתם שתיארתם? נמקו.

לדוגמה: עבור קבוצת המספרים $\{1, 3, 6, 7, 10, 12, 15, 16, 20\}$ ו- $k=3$, נחלק לקבוצות הבאות: $\{1, 3, 6\}$, $\{7, 10, 12\}$, $\{15, 16, 20\}$.

מועד ג'

10 מתוך 10

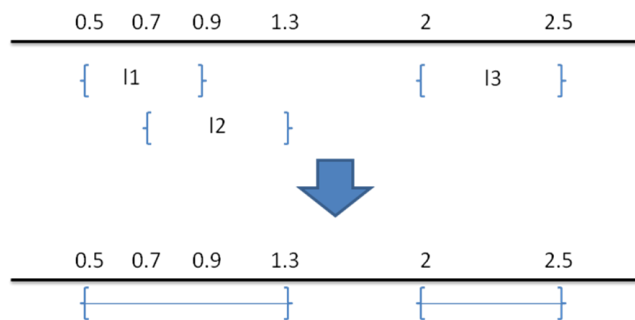
מספר תעודת זהות:

מספר מחברת:

ב. בהינתן n מספרים מתחום סדור, ומס' k כך ש- $k < n$, תאר אלגוריתם המוצא בזמן $O(n)$ את k המספרים הקרובים ביותר בערכם לחציון?

ג. נתונים n קטעים על הישר. תארו אלגוריתם המוצא מהו אורך איחוד הקטעים. לדוגמה

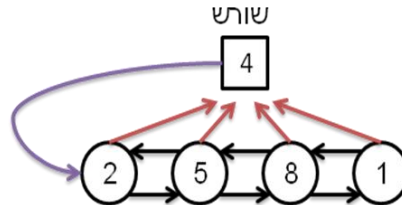
עבור הקטעים $[0.5, 0.9]$, $[0.7, 1.3]$, $[2, 2.5]$ נקבל:



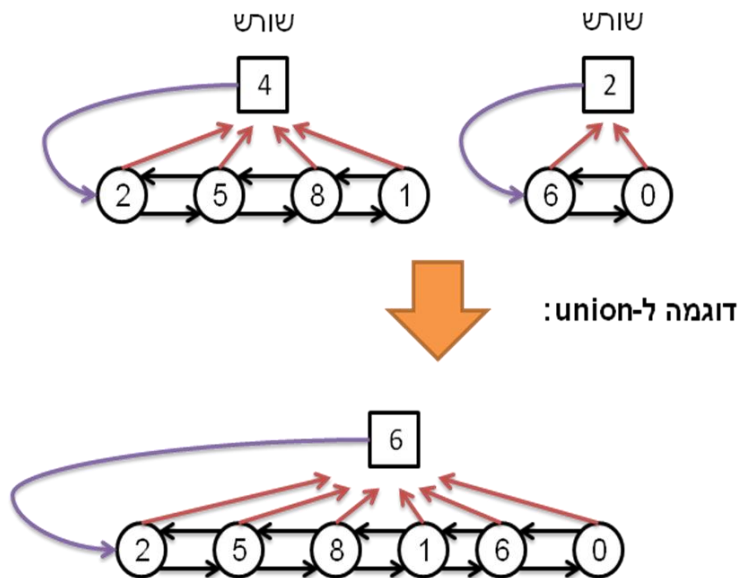
אורך איחוד הקטעים: $0.8 + 0.5 = 1.3$

שאלה 6 (14 נקודות)

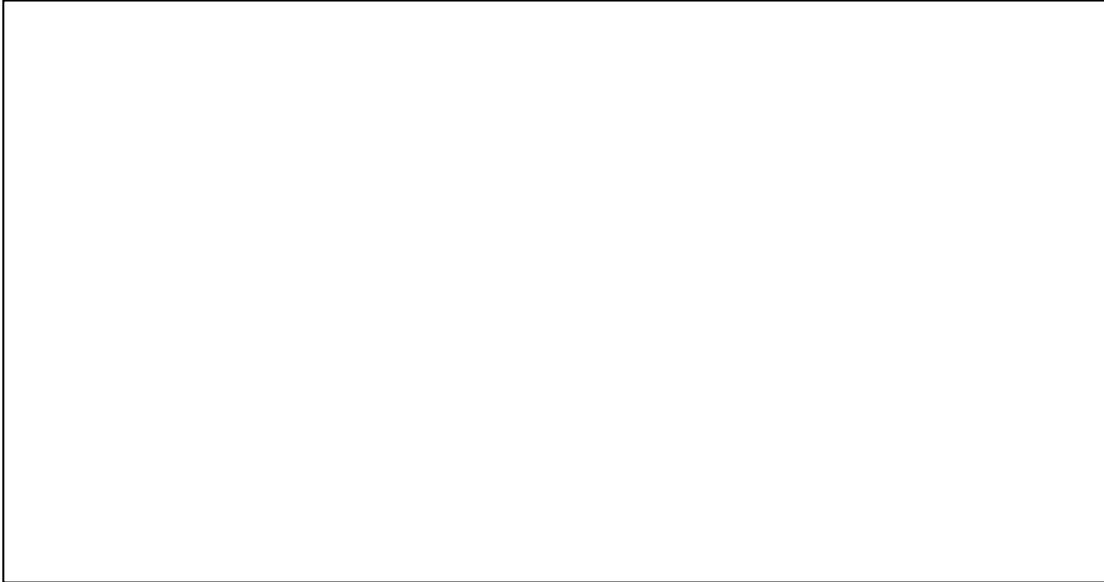
להלן מבנה נתונים הפותר את בעיית ה-union find. נחזיק קבוצה ע"י רשימה משורשרת דו-כיוונית של אבריה כאשר כל אלמנט מצביע לצומת מיוחד המייצג את הקבוצה (נכנה אותו שורש הקבוצה). שורש הקבוצה מצביע לרשימה המשורשרת וגם מכיל את מספר האלמנטים בקבוצה. לדוגמה הנה ייצוג של הקבוצה {1,5,8,2}, גודל הקבוצה (4) נשמר בשורש.



- פעולת ה-find מקבלת מצביע לאלמנט ומחזירה את המצביע לשורש שמאוכסן בצומת המייצג את האלמנט.
- פעולת ה-union מוסיפה את אברי הקבוצה הקטנה בזה אחר זה לרשימה של הקבוצה הגדולה ומעדכנת את המצביע לשורש באברים אלו, שיצביע לשורש בקבוצה הגדולה. בנוסף, הפעולה מעדכנת את מספר האברים בשורש הקבוצה הגדולה, ומוחקת את שורש הקבוצה הקטנה.

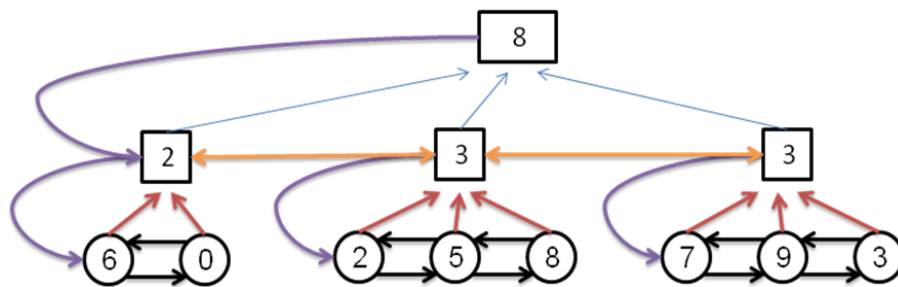


א. כמה זמן במקרה הגרוע ייקח לבצע סדרה בת m פעולות כאשר מתחילים מאוסף של n קבוצות, כל אחת מכילה רק אבר אחד? הנח $m > n$. הוכח את נכונות תשובתך.



ב. נשנה את מבנה הנתונים באופן הבא. נסמן $k = \lfloor \log n \rfloor$. קבוצות שגודלן k לכל היותר נייצג כמו קודם. קבוצה S בגודל k תיוצג ע"י אוסף של k קבוצות בגודל k בדיוק (נקרא להן **מלאות**), וקבוצה אחת לכל היותר בגודל k (חסרה). תתי קבוצות אלו מיוצגות כמו שתואר בחלק א' של השאלה. שורשי תתי הקבוצות מסודרים ברשימה משורשרת דו-כיוונית ומצביעים לצומת חדש המייצג את הקבוצה כולה. שורש זה מכיל מצביע לרשימת שורשי תתי הקבוצות ואת גודל הקבוצה. הקבוצה החסרה (אם קיימת) תהיה תמיד ראשונה ברשימה.

דוגמה לקבוצה גדולה עם שתי קב' מלאות ואחת חסרה, $k = \log n = 3$



נבצע את הפעולות באופן הבא:

- נבצע **find** ע"י מציאת שורש תת הקבוצה המאוכסן בצומת המייצג את האיבר ומשם נמשיך לשורש הקבוצה (אם היא קב' גדולה).
- בכדי לבצע פעולת **union** של תת קבוצות, נשתמש בפעולת ה-**union** כמו שהוגדרה קודם, עם השינוי הבא: במידה וגודל שתי התת קבוצות יחדיו הוא יותר מ- k אזי ה-**union** יחזיר שתי תת קבוצות, אחת מלאה ואחת חסרה. **union** זה מתבצע כמו קודם רק שאין הוא מעתיק את כל איברי הקבוצה הקטנה במידה ואין יותר מקום בקבוצה הגדולה.
- **union** של קבוצות נבצע באופן הבא:
 - אם שתי הקבוצות קטנות וגודלן יחדיו קטן או שווה ל- k אזי הפעולה תבוצע כאיחוד תתי קבוצות כמו קודם.
 - אם שתי הקבוצות קטנות וגודלן יחדיו גדול מ- k אזי נאחד את תתי הקבוצות ונקבל תת קבוצה אחת מלאה ואחת חסרה. ניצור קבוצה חדשה שאלו הן תתי הקבוצות שלה ונחזיר אותה.
 - אם קב' אחת קטנה ואחת גדולה, נאחד את הקבוצה הקטנה עם תת הקבוצה החסרה של הקבוצה הגדולה. נוסיף את תתי הקבוצות המתקבלות לרשימת תתי הקבוצות של הקבוצה הגדולה. נחזיר את הקבוצה הגדולה.
 - אם שתי הקבוצות גדולות נעביר את המלאות מן הקבוצה הקטנה יותר לרשימת תת הקבוצות בקב' הגדולה יותר. כמו כן נאחד את החסרות ונוסיף את תתי הקבוצות המתקבלות לרשימת תתי הקבוצות של הקבוצה הגדולה. נחזיר את הקבוצה הגדולה.

הוכח כי זמן הריצה של סדרה בת m פעולות כאשר מתחילים מאוסף של n קבוצות, כל אחת

מכילה רק אבר אחד הוא $O(n \log \log n + m)$.

נספח לשאלה 2: חמשת מקרי ה-insert של עץ אדום שחור

