

מספר קורס: 0368.2158

סמסטר א', מועד א', תשע"א

17.2.2011

## מבחן במבני נתונים

**פרופ' חנוך לוי, ד"ר ערן הלפרין, ירון אורנשטיין ויהב נוסבאום**

משך המבחן שלוש שעות

הבחינה עם חומר סגור, ללא מחשבים/מחשבוניס. מותר להביא דף עזר אחד A4 זו-צדדי.

### הוראות כלליות:

1. יש לרשום מספר ת.ז. ומספר מחברת בראש כל דף.
2. המחברות הן לטיוטה בלבד ולא תיבדקנה.
3. הסברים לשאלות האמריקאיות – **חובה!!!**
4. **תשובות סופיות לשאלות האמריקאיות יש לכתוב על הטופס המצורף.**
5. בכל שאלה אמריקאית יש תשובה **אחת נכונה ביותר**. רק תשובה זו תחשב לנכונה.
6. מותר להשתמש באלגוריתמים שנלמדו בכיתה כקופסאות שחורות.
7. בכל מקום בו מוזכרים חסמי זמן הכוונה היא למקרה הגרוע ביותר, אלא אם כן צוין אחרת במפורש.
8. **אין** לכתוב אלגוריתמים בפסאודו-קוד אלא להסביר (באופן משכנע) את הרעיון הכללי.
9. כתבו תשובות קצרות ומדויקות. תשובות לא ממוקדות, ארוכות, מסובכות או מסורבלות יקבלו ניקוד חלקי גם אם הן נכונות.
10. גם אם שאלה נראית לכם דומה לשאלה מבחינות קודמות, היא לא בהכרח זהה.

**הצלחה!!!**

**שאלה 1 (10 נקודות):**

נתונה נוסחת הנסיגה הבאה:

$$T(n) = \begin{cases} 8T(\lfloor n/4 \rfloor) + n, & n > 1 \\ 1 & n = 1 \end{cases}$$

סטודנט טען שהנוסחה מקיימת  $T(n) = O(n)$  וסיפק את ההוכחה הבאה:i. עבור  $n=1$  הטענה מתקיימת מיידית.ii. עבור  $n > 1$  נוכיח באינדוקציה.a. נניח  $T(i) \leq ci$  לכל  $i \leq n-1$ b. נציב את נוסחת הנסיגה בהנחה (a) ונקבל:  $T(n) \leq 8(cn/4) + n$ c. מ-b נקבל  $T(n) = O(n)$ .בחר את התשובה הנכונה ביותר:

א. הסטודנט צדק בטענה ובהוכחה.

ב. הסטודנט טעה בטענה ובהוכחה. החלק הלא נכון בהוכחה הינו (ii.c).

ג. הסטודנט טעה בטענה ובהוכחה. החלק הלא נכון בהוכחה הינו (i).

ד. הסטודנט צדק בטענה וטעה בהוכחה.

ה. הטענה אינה נכונה אך אם נחליף בנוסחת הנסיגה את המספר 8 ב-4 הטענה תהיה נכונה.

ו. תשובות ב' ו-ג' נכונות.

נימוק קצר: (חובה)

ב+ג נכון (10 נקודות). ו קבלה נקוד + בונוס (13 נקודות סה"כ).

**שאלה 2 (10 נקודות):**

הגדירו מבנה נתונים הכולל את הפעולות Insert ו-Delete-Min כמו בערמה, ובנוסף פעולה חדשה *Increase-Large* – פעולה זאת תגדיל ב-1 את ערכם של  $n/2$  האיברים הגדולים מתוך  $n$  האיברים שבערמה. זמני הריצה worst case לפעולה הטובים ביותר (מבין האפשרויות המוצעות) שבהם אפשר לממש את מבנה הנתונים הם:

- א. Insert ב- $O(\log n)$ , Delete-Min ב- $O(\log n)$ , Increase-Large ב- $O(1)$ .
- ב. Insert ב- $O(\log n)$ , Delete-Min ב- $O(1)$ , Increase-Large ב- $O(\log n)$ .
- ג. Insert ב- $O(1)$ , Delete-Min ב- $O(1)$ , Increase-Large ב- $O(\log n)$ .
- ד. Insert ב- $O(\log n)$ , Delete-Min ב- $O(\log n)$ , Increase-Large ב- $O(n)$ .
- ה. Insert ב- $O(1)$ , Delete-Min ב- $O(\log n)$ , Increase-Large ב- $O(\log n)$ .
- ו. Insert ב- $O(\log n)$ , Delete-Min ב- $O(\log n)$ , Increase-Large ב- $O(\log n)$ .

נימוק קצר והסבר המבנה (חובה):

Answer is (1): Insert and delete\_min in  $O(\log n)$  and increase\_large in  $O(1)$ . You keep two heaps – min-max for the  $n/2$  smallest numbers and min for the  $n/2$  largest numbers. You also keep a variable for the offset. Increase-large increases the offset by 1. When delete-min or insert are performed one has to move elements from one heap to another and update their values according to the offset.

**שאלה 3 (9 נקודות):**

כזכור, פעולה בסיסית במודל ההשוואות מורכבת מהשוואה בין שני איברים. אנחנו בוחנים חלופה של מודל ההשוואות, שבה פעולה בסיסית מורכבת ממיון  $k$  איברים בבת אחת. מה הוא החסם התחתון ההדוק ביותר שניתן למצוא על פי המודל החלופי למיון  $N$  איברים?

א.  $\Omega(N)$

ב.  $\Omega(k \log k)$

ג.  $\Omega(N \log N)$

ד.  $\Omega\left(\frac{N \log N}{k}\right)$

ה.  $\Omega\left(\frac{N \log N}{\log k}\right)$

ו.  $\Omega\left(\frac{N \log N}{k \log k}\right)$

נימוק קצר: (חובה)

Answer is (5) Theta ( $N \log N/k \log k$ ). When considering the decision tree in the lower-bound proof, each vertex has  $k!$  children and therefore the height of the tree is at least  $\log(n!)/\log(k!)$  which gives the results.

**שאלה 4 (9 נקודות):**

מימשנו עץ חיפוש בינארי (לא בהכרח מאוזן) על  $n$  איברים. נסמן ב- $k$  את מספר האיברים בעץ שפעולת `find` עבורם לוקחת זמן קבוע (כלומר, שאינו תלוי ב- $n$ ). מהו  $k$ ?

א.  $k = 1$

ב.  $k = O(1)$

ג.  $k = O(\log n)$

ד.  $k = \lfloor n/2 \rfloor$

ה.  $k = O(n)$

ו. לא ניתן לדעת.

נימוק קצר: (חובה)

תשובה ב: לא יתכן שבעומק קבוע יש יותר ממספר קבוע של צמתים כי לכל צומת יש 2 בנים לכל היותר

**שאלה 5 (10 נקודות):**

אנחנו רוצים לצבוע את הצמתים של עץ חיפוש בינארי מלא (כלומר, לכל צומת שני בנים) באדום ובשחור, כך שהצביעה תקיים את הכללים של עץ אדום-שחור. איזה מהתנאים הבאים הוא תנאי מספיק לכך שקיימת צביעה שכזאת?

- א. כל העלים נמצאים באותו מרחק מהשורש.
- ב. המרחק הגדול ביותר משורש לעלה הוא לכל היותר פי שתיים מהמרחק הקטן ביותר משורש לעלה.
- ג. עומק תת-העץ הימני של השורש הוא לכל היותר פי שתיים מעומק תת-העץ השמאלי של השורש.
- ד. תשובות א' ו-ב' נכונות.
- ה. תשובות א' ו-ג' נכונות.
- ו. אף תשובה אינה נכונה.

נימוק קצר: (חובה)

תשובה א: נצבע את כל העץ בשחור. ניתן לבנות עץ שסותר את ב ולכן ד אינה נכונה

**שאלה 6 (9 נקודות):**

נתונים לנו שני עצי  $+2-4$ , האחד בגובה  $h_1$  והשני בגובה  $h_2$ . הגבהים  $h_1, h_2$  ידועים לנו, כמו כן ידוע לנו ש- $h_1 > h_2$ , ושכל האיברים בעץ הראשון (שגובהו  $h_1$ ) קטנים מכל האיברים שבעץ השני. אנחנו רוצים למזג את שני העצים לאחד. באיזה זמן ריצה ניתן לעשות זאת? בחרו את התשובה ההדוקה ביותר.

א.  $O(h_1)$ .ב.  $O(h_2)$ .ג.  $O(h_1 h_2)$ .ד.  $O(h_1/h_2)$ .ה.  $O(h_1+h_2)$ .ו.  $O(h_1-h_2)$ .

נימוק קצר: (חובה)

תשובה ו: נתלה את העץ הקטן על המסלול הימני ביותר של העץ הגדול ונתקן מעלה.

**שאלה 7 (10 נקודות):**

בשאלה הזאת נדון בעץ בעל  $n > 100$  צמתים שדרגת השורש שלו היא 2. נגדיר את חוסר האיזון של העץ כיחס בין עומק תת-העץ השמאלי של השורש לעומק תת-העץ הימני של השורש (יחס = המספר הראשון חלקי המספר השני). באיזה מן העצים הבאים נוכל לקבל את חוסר האיזון הגדול ביותר:

- א. עץ שמייצג עץ אדום-שחור.
- ב. עץ שמייצג עץ 2-4 ודרגת השורש שלו היא 2.
- ג. עץ שמייצג מבנה union-find (עם union by rank) ודרגת השורש שלו היא 2.
- ד. עץ שמייצג מבנה union-find ללא union by rank ודרגת השורש שלו היא 2.
- ה. תשובות א' ו-ב' נכונות.
- ו. תשובות ג' ו-ד' נכונות.

נימוק קצר: (חובה)

תשובה ד: זה העץ היחיד שבו אין איזון של העומק



**שאלה 8 (9 נקודות):**

יהי  $A$  מערך של  $N$  מספרים. נסמן המספר במקום ה- $i$  הוא  $A_i$ . נאמר שיש הפך בין  $A_i, A_j$  אם  $i < j$  ו- $A_i > A_j$ . נסמן ב- $I$  את מספר ההפכים במערך.

מה זמן הריצה של האלגוריתם היעיל ביותר, המקבל את  $A$  ומחזיר כמה הפכים יש בו, כלומר את  $I$ ?

בחרו את החסם העליון ההדוק ביותר עבור זמן הריצה של האלגוריתם:

- א.  $O(I)$
- ב.  $O(I+N)$
- ג.  $O(N \log N)$
- ד.  $O(\min\{I+N, N \log N\})$
- ה.  $O(N + N \log (I/N))$
- ו. אף תשובה אינה נכונה.

נימוק קצר: (חובה)

פתרון ה'. בהרצאה ראינו מיון הכנסה באמצעות FINGER TREE. זמן הריצה שלה"כ  $O(N + N \log (I/N))$  ולכל הכנסה  $O(\log(Ik))$ , כש- $Ik$  הוא מספר ההפכים עד איבר ה- $K$ .

צריך לשים לב לשני דברים: התיקונים בשביל עץ מאוזן בעץ אדום שחור עולים שלה"כ זמן לינארי, ובעלות החיפוש כבר התחשבנו. אפשר לבצע עדכון שדה SIZE בזמן העלייה בחיפוש (בעזרת זכירת חובות עדכון), ולכן זה לא מוסיף לעלות.

**שאלה 9 (12 נקודות):**

האלגוריתם "select" הדטרמיניסטי, פועל באופן הבא:

- המערך מחולק לקבוצות בגודל 5.
- ממיינים כל קבוצה.
- בוחרים pivot בתור החציון של החציונים.
- מבצעים "Partition" עם ה-pivot שנבחר.

**הניחו שבמקום קבוצות בגודל 5 אנו לוקחים קבוצות בגודל  $\log(n)$ ? מה יהיה זמן הריצה של האלגוריתם:**

א. (2 נק') מהי עלות צעד b?

ב. (2 נק') מהי עלות צעד c?

ג. (8 נק') מהי העלות הכוללת של האלגוריתם?

- Cost = Theta ( $n \log \log n$ ): Explanation: Each set of size  $\log n$  is sorted in Theta ( $\log n * \log \log n$ ) time. There are  $n/\log n$  such sets.
- Cost =  $T(n/\log n)$ , where  $T(x)$  is the running time of the algorithm on an array of length  $x$ . Explanation: A recursive call.
- Cost:  $O(n \log \log n)$ . Explanation: It can be shown that the pivot is larger (and smaller) than at least  $n/4$  elements, and therefore the recurrence formula is  $T(n) \leq C*n \log \log n + T(n/\log n) + T(3n/4)$ . By induction it can be shown that this results in  $T(n) = \text{Theta}(n \log \log n)$ .

בבדיקה נרשמו הקודים לפתרון והנה הניקוד שלהם וההסבר שלהם:

- A עם כמה עולה למיין קבוצה אחת (א')
- B (ג') עלות כוללת לינארית, כי האיבר החופשי לינארי
- C יש רק נוסחת נסיגה, לא פתורה (ג')
- D (ג) נוסחת נסיגה שגויה בחלק של ה-3/4
- E חסר חסם תחתון (ג)
- F טעות בפתרון נוסחת נסיגה, לא חסום ע"י ( $c n \log \log n$ ) בהוכחה באינדוקציה שלב אחרון
- G פתרון (ב) עם O במקום
- H פתרון שגוי של (ג)
- I טעות ב(ג) כתוצאה מטעות נגררת
- J פתרון שגוי (ג) של  $n \log n \log \log n$
- K פתרון זמן קבוע לסעיף (ב)
- L פתרון שגוי של (ב)
- M פתרון של (ג) לפי שיטת המאסטר
- N נוסחת נסיגה נכונה, פתרון שגוי של נוסחת נסיגה (ג)
- O פתרון שגוי של (א)
- P הסבר לא מספק ב-ג)

-3	-2	-3	-3	-2	-2	-2	-3	-7	0	-3	0	-3	-3	-3	0
P	O	N	M	L	K	J	I	H	G	F	E	D	C	B	A

**שאלה 10 (12 נקודות):**

**תזכורת:** ב-perfect hash אנחנו בוחרים פונקציית hash אוניברסלית אקראית בכדי למפות מרחב נתון  $U$  בגודל  $n$  למערך בגודל  $n$ . יהי  $n_i$  מספר האיברים שמופו לתא ה- $i$ . לאחר מכן, משתמשים בפונקציית hash אוניברסלית אקראית בכדי למפות את האיברים שמופו לתא ה- $i$  למערך אחר בגודל  $n_i^2$ .

**שאלה:**

**נגדיר:** משפחה כמעט אוניברסלית של פונקציות היא משפחה של פונקציות ממרחב  $U$  ל- $\{0, \dots, m-1\}$  כך שכל זוג מפתחות  $k_1, k_2$  ב- $U$  מתקיים  $\Pr(h[k_1] = h[k_2]) \leq 1/\sqrt{m}$ . נניח כי אנו מקבלים משפחה כמעט אוניברסלית כאשר  $|U|=n$ . ניקח פונקציה אקראית מהמשפחה ונשתמש בה כדי למפות את האלמנטים של  $U$  למערך בגודל  $m$ .

א. (2 נק') מהו החסם העליון הטוב ביותר על תוחלת מספר ההתנגשויות אם  $m=n^4$ ?

ב. (2 נק') מהו החסם העליון הטוב ביותר על תוחלת מספר ההתנגשויות אם  $m=n^2$ ?

ג. (8 נק') מהי תוחלת זמן הריצה הטובה ביותר שאתם יכולים להשיג כדי לבנות perfect hash?

1. Answer = 0.5. There are  $n^2/2$  possible collisions and the probability for each is  $1/\sqrt{m} = 1/n^2$ , so by linearity of expectation we get 0.5.
2. Answer =  $n/2$ . There are  $n^2/2$  possible collisions and the probability for each is  $1/\sqrt{m} = 1/n$ , so by linearity of expectation we get  $n/2$ .
3. Answer =  $O(n^2)$ . Two layer hashing – the first layer has  $n^2$  cells and so we can find in quadratic time a function that maps the elements of  $U$  with  $< n$  collisions. Then, for each cell  $i$ , we can map its elements to an array of length  $n_i^2$ , so that in time  $O(n_i^2)$  we will find a hash function that maps to the array with no collisions. Now,

$$\sum n_i^4 \leq \left(\sum n_i^2\right)^2 \leq n^2.$$

# סוף