אוניברסיטת ת"א
סמסטר ב', תשפ"ה
הגשה עד ה-11.5 ב-moodle

תרגיל בית 3 במבני נתונים

על כל התשובות להיות מנומקות. בכל שאלה יש לבחור במימוש היעיל ביותר האפשרי מבחינת סיבוכיות זמן. יש לענות על השאלות במקומות המוגדרים לכך.

<u>שאלה 1</u>

א. תארו מבנה נתונים התומך בפעולות הבאות על קבוצה S מתחום בעל סדר מלא.

- Median(S): מחזיר את החציון ב-S (אם מס' האיברים זוגי – חציון תחתון, כלומר האיבר שימוקם במקום ה-n/2 אם נמיין את האיברים)
- Min(S): מחזיר את האיבר הקטן ביותר ב-S
- Max(S): מחזיר את האיבר הגדול ביותר ב-S
- Insert(x,S) : מוסיף איבר x ל-S
- Delete(x,S) : מניחים ש-x נמצא לפני הפעולה ב-S. הפעולה מוציאה את x מ-S.

על הפעולות median, min, max לקחת $O(1)$ זמן במקרה הגרוע, ועל הפעולות insert ו-delete לקחת $O(\log n)$ זמן במקרה הגרוע. כמו כן תארו בקצרה כיצד לבצע כל פעולה.

ב. שנו במידת הצורך את מבנה הנתונים שהגדרתם בסעיף הקודם כך שנוכל גם למצוא את האלמנט ה- i בגודלו בזמן $O(\log(i \; mod \; \frac{n}{2}))$, לדוגמה עבור $i = \frac{n}{2} + 5$ העלות של הפעולה צריכה להיות $O(\log(5)) = O(1)$. תארו במדויק כיצד לבצע פעולה זו.

שאלה 2

תכננו מבנה נתונים המכיל מפתחות טבעיים ללא חזרות (כלומר המפתחות ייחודיים) ותומך בפעולות Insert,Delete,Search בזמן $O(\log n)$ וכן בפעולה IncreaseAll(k), אשר מוסיפה את הטבעי k לכל המפתחות, בזמן $O(1)$. בנוסף, על מבנה הנתונים לתמוך בפעולה $\mathrm{EvenLess}(x)$ שמחזירה בזמן $O(\log n)$ את סכום המפתחות הזוגיים במבנה שערכם לכל היותר x (נניח ש-x נמצא במבנה), כאשר n הוא מספר המפתחות במבנה.

שאלה 3

נתון עץ AVL בו מאוחסנים מפתחות טבעיים שונים זה מזה. בכל צומת v בעץ שמור size(v), גודל תת העץ של v.

א. תארו אלגוריתם המוצא את הטבעי המינימלי שאינו שייך לעץ בזמן $O(log\ n)$.
רמז: חישבו מהי משמעות הביטוי $k - rank(k)$, כאשר $rank(k)$ מציין את מספר המפתחות בעץ הקטנים או שווים ל-k.

ב. נכליל את הסעיף הקודם - נסמן את קבוצת המפתחות השמורים בעץ ב-S, כאשר $|S| = n$.
נגדיר: nextMissingAfter(i) - המספר הטבעי המינימלי $j \notin S$ הגדול מ-i .
למשל, עבור עץ שקבוצת המפתחות השמורים בו היא S = {11,10,9,6,5,4,2,1}, מתקיים NextMissingAfter(3) = 7, NextMissingAfter(9) = 12, NextMissingAfter(5) = 7.
הראו כיצד ניתן לחשב את הפונקציה nextMissingAfter(i) בסיבוכיות זמן $O(log\ n)$.

<u>שאלה 4</u>

א. הוכיחו כי בעץ AVL בגובה h, כל העלים בעומק לפחות $h/2$.

ב. הוכיחו כי כל סדרה בת n הכנסות לעץ AVL, גם הטובה ביותר וגם הגרועה ביותר, היא בעלות $\theta(n\log n)$ (הכנסות עם חיפוש שמתחיל מהשורש).

ג. בתרגול ניתחנו מצב בו בהינתן m ועץ AVL מתבוננים בתת העץ המינימלי שמכיל את m המפתחות הקטנים ביותר. מצאו חסם עליון אסימפטוטי של גודל תת-עץ זה כפונקציה של m.
תזכורת: בהרצאה הוכחתם כי בעץ AVL בעל n צמתים בגובה h מתקיים ש- $h \leq \log_{\phi} n$.

<u>שאלה 5</u>

הציעו מימוש למבנה נתונים התומך בפעולות הבאות, על ישרים מהצורה $y = ax + b$:

- Search(a,b) – האם הישר $y = ax + b$ נמצא במבנה?
- Insert(a,b) – הכנסת הישר $y = ax + b$ למבנה, אם אינו חותך בקטע $[0,1]$ אף ישר אחר שנמצא כבר במבנה. כלומר הישר החדש ייכנס אם אין שום ישר אחר במבנה שנחתך איתו בנקודה (x_1, y_1) המקיימת $0 \le x_1 \le 1$. אם תנאי זה לא מתקיים הפעולה לא תבצע דבר.

  סיבוכיות הזמן הדרושה עבור שתי הפעולות היא $O(\log n)$ במקרה הגרוע, כאשר n הוא מספר הישרים במבנה. תארו תחילה מה כולל המבנה שלכם (כלומר איזה מידע נשמר וכיצד) ולאחר מכן את הפעולות השונות, כולל הסבר קצר מדוע הן עומדות בדרישות הסיבוכיות.

<u>רמז</u>: ניתן לייצג ישר ע"י שתי נקודות במישור.

שאלה 6

בשאלה הזו נוכיח טענה שראיתם בהרצאה, לפיה מיון הכנסה הממומש בעץ עם מצביע למקסימום עובד בזמן $O(n\log\left(\frac{I}{n}\right))$, כאשר I הוא מספר ההיפוכים בקלט. בביטוי הזה יש הנחה כי $I \geq n$ (אחרת הלוג מקבל ערך שלילי). לשם פשטות, וכדי להיות כלליים יותר ולוותר על הנחה זו, נחליף את הביטוי ב $O(n\log\left(\frac{I}{n}+2\right))$.

נתון מערך A בגודל n של איברים מתחום סדור כלשהו. נגדיר היפוך בתור זוג אינדקסים $i<j$ כך ש-$A[i]>A[j]$, ונסמן את מספר ההיפוכים במערך ב-I. ניתן לשים לב כי באופן כללי $0 \leq I \leq \binom{n}{2}$ וככל שיש פחות היפוכים כך המערך קרוב יותר לממוין.

נראה כיצד ניתן למיין את A בזמן $O(n\log\left(\frac{I}{n}+2\right))$ בעזרת שימוש ב-AVL Finger Tree עם מצביע למקסימום. המיון יתבצע ע"י הכנסת האיברים לעץ וסריקת in-order, כאשר בעת ההכנסה החיפושים יתחילו מהצומת המקסימלי בעץ, כפי שמימשתם בתרגיל המעשי.

א. חסמו מלמעלה את סך עלויות התיקון של העץ לאורך סדרת ההכנסות.

ב. נסמן ב-d_i את מספר האיברים לפני האינדקס i שגדולים ממנו. חסמו את עלות החיפוש בעת הכנסת האיבר ה-i כפונקציה של d_i והסיקו כי סך עלויות החיפוש לאורך סדרת ההכנסות הוא $O(\log\Pi_{i=1}^{n}(d_i+2))$.

ג. השתמשו באי-שוויון הממוצעים, והסיקו את הנדרש.