

## מבני נתונים – פרויקט חקר תכנותי מספר 1 – עץ דרגות

### הקדמה:

בתרגיל זה שני חלקים:

1. חלק המעשי: מימוש של List באמצעות עץ AVL. עמודים 1-2 במסמך זה מתארים את החלק הזה.
2. חלק ניסוי-תיאורטי: בהתבסס על המימוש מהחלק המעשי, נבצע מספר "ניסויים" עם ניתוח תיאורטי נלווה ושאלות חקר. עמודים 3-5 מתארים את החלק הזה. שימו לב: בסוף המסמך (עמוד 5) ישנן הוראות הגשה – הקפידו לפעול לפיהן. תאריך הגשה: 30/12. בנוסף, יש לעקוב אחר השרשור הנעוץ בפורום בו נפרסם הבהרות חשובות.

### חלק מעשי

#### דרישות

בתרגיל זה נממש את ה ADT רשימה באמצעות עץ AVL. לכל איבר ברשימה יש ערך (info). המימוש יהיה בשפת **python** 3.9 וצריך להיות מבוסס על קובץ השלד המופיע באתר הקורס. הפעולות שיש לממש הן:

פעולה	תיאור
<code>empty()</code>	הפונקציה מחזירה ערך <code>TRUE</code> אם ורק אם הרשימה ריקה.
<code>retrieve(i)</code>	הפונקציה מחזירה את ערך האיבר במקום ה- <code>i</code> אם קיים, אחרת היא מחזירה <code>None</code> .
<code>insert(i, s)</code>	הכנסת איבר בעל ערך <code>s</code> לרשימה במקום ה- <code>i</code> , במידה וקיימים לפחות <code>i</code> איברים ברשימה. הפונקציה מחזירה את מספר פעולות האיזון שנדרשו בסה"כ בשלב תיקון העץ על מנת לשמר את תכונת האיזון.
<code>delete(i)</code>	מחיקת האיבר במקום ה- <code>i</code> ברשימה, אם הוא קיים. הפונקציה מחזירה את מספר פעולות האיזון שנדרשו בסך הכל בשלב תיקון העץ על מנת לשמר את תכונת האיזון. אם לא קיימים מספיק איברים ברשימה הפונקציה מחזירה <code>-1</code> .
<code>first()</code>	מחזירה את ערך האיבר הראשון ברשימה, או <code>None</code> ברשימה ריקה.
<code>last()</code>	מחזירה את ערך האיבר האחרון ברשימה, או <code>None</code> ברשימה ריקה.
<code>listToArray()</code>	הפונקציה מחזירה מערך המכיל את איברי הרשימה לפי סדר האינדקסים, או מערך ריק אם הרשימה ריקה.
<code>length()</code>	הפונקציה מחזירה את מספר האיברים ברשימה.
<code>permutation()</code>	הפונקציה מחזירה את אותם האיברים ברשימה בסדר אקראי.
<code>sort()</code>	הפונקציה מסדרת את איברי הרשימה בסדר עולה.
<code>concat(lst)</code>	הפונקציה מקבלת רשימה. על הפונקציה לשרשר אותה אל סוף הרשימה הנוכחית. על הפעולה לרוץ בזמן $O(\log n)$ . יש להחזיר את הערך המוחלט של הפרש הגבהים של עצי AVL שמוזגו.
<code>search(val)</code>	החזרת האינדקס הראשון ברשימה בו מופיע הערך <code>val</code> , או <code>-1</code> אם לא קיים כזה.

בנוסף למימוש הפונקציות האלו, יש לממש את מחלקת **AVLNode** כפי שמתואר בקובץ. מטעמי נוחות, נדרוש שלכל עלה יהיו 2 בנים "וירטואליים", כלומר, צמתים שלא מייצגים איברים במבנה הנתונים. באופן זה, נוח יותר לממש גלגולים מכיוון שלכל צומת יהיו 2 בנים זה חוסך טיפול במקרי קצה.

למחלקה AVLNode יש את המתודות הבאות (המפרט המלא נמצא בקובץ השלד):  
getHeight – מחזיר את הגובה של הצומת, או 1 – אם הצומת הוא וירטואלי.  
getValue – מחזיר את הinfo של הצומת או None אם הצומת הוא וירטואלי.  
getLeft – מחזיר את הבן השמאלי של הצומת, או None אם אין כזה.  
getRight – מחזיר את הבן הימני של הצומת, או None אם אין כזה.  
getParent – מחזיר את ההורה של הצומת, או None אם אין כזה.  
isRealNode – מחזיר TRUE אם הצומת מייצג צומת אמיתי בעץ (קרי: צומת שאינו וירטואלי).

### הערות חשובות:

1. המימוש יבוצע על ידי מילוי קובץ השלד. במידת הצורך, ניתן להרחיב את המימוש (למשל להוסיף פונקציות עזר שאינן מופיעות בשלד), אך אסור לשנות את הגדרות הפונקציות לעיל. על כל הפונקציות/מחלקות להופיע בקובץ יחיד.
2. אין להשתמש באף מימוש ספרייה של מבנה נתונים.

### סיבוכיות

יש לתעד בקוד ובמסמך נפרד (ביותר פירוט) את סיבוכיות זמן הריצה במקרה הגרוע (האסימפטוטית, במונחי O הדוקים) של כל פונקציה שמכילה לולאות/רקורסיה, כתלות במספר האיברים בעץ n. עליכם להשיג סיבוכיות זמן ריצה (במקרה הגרוע ביותר) נמוכה ככל הניתן עבור כל אחת מהפונקציות.

### פלט

אין צורך בפלט למשתמש.

### תיעוד

בנוסף לבדיקות אוטומטיות של הקוד שיוגש, קובץ המקור ייבדק גם באופן ידני. חשוב להקפיד על תיעוד לכל פונקציה, וכמות סבירה של הערות. הקוד צריך להיות קריא, בפרט הקפידו על בחירת שמות משתנים ועל אורך השורות.  
יש להגיש בנוסף לקוד גם מסמך תיעוד חיצוני. המסמך יכלול את תיאור המחלקה שמומשה, ואת תפקידו של כל חבר במחלקה. עבור כל פונקציה במחלקה יש לפרט מה היא עושה, כיצד היא פועלת ומה סיבוכיות זמן הריצה שלה. בפרט, אם פונקציה קוראת לפונקציית עזר, יש להתייחס גם לפונקציית העזר בניתוח. עבור פונקציות שעולות זמן קבוע יספיק להביא רק תיאור קצר ולא לפרט את ניתוח הסיבוכיות.

### בדיקות

התרגילים ייבדקו באמצעות תוכנת טסטר שקוראת לפונקציות המפורטות מעלה בתרחישים שונים, ומוודאת את נכונות התוצאות. קובץ הטסטר שלנו לא יפורסם לפני הבדיקות. מומלץ מאוד לממש אוסף בדיקות עבור המימוש, לא בשביל ההגשה, אלא כדי לבדוק שהקוד לא רק רץ, אלא גם נכון!

בקובץ שתגישו לא תהיה פונקציית **main** ולא יהיו הרצות קוד/הדפסות, דבר זה יפגע בטסטר שיבדוק לכם את התרגילים.

## חלק ניסויי/תיאורטי

### שאלה 1:

בשאלה זאת נדון בכמות צעדי האיזון הממוצעת הנדרשת בסדרת הכנסות, בסדרת מחיקות ובסדרה מעורבת של הכנסות ומחיקות.

- לצורך הניתוח, נבנה עצי AVL בגדלים שונים. מספר איברים שנכניס לעץ יהיה  $n = 2^i \cdot 1500$  כאשר  $i=1, \dots, 10$ . כלומר, עבור  $i=1$  העץ בגודל 3000, ועבור  $i=10$  העץ בגודל כמיליון וחצי.
- לכל גודל של עץ, נבצע 3 ניסויים נפרדים:
  - נכניס איברים בסדר אקראי.
  - נכניס איברים בסדר אקראי (לא נספר), ולאחר מכן נמחק אותם בסדר אקראי.
  - נכניס  $n/2$  איברים בסדר אקראי, לאחר מכן נבצע  $n/4$  הכנסות ומחיקות אקראיות לסירוגין.

1. עבור כל ניסוי, יש לציין את מספר פעולות האיזון שנדרשו כדי לתקן את העץ עבור הפעולות שהודגשו.

מספר סידורי $i$	ניסוי 1 - הכנסות	ניסוי 2 - מחיקות	ניסוי 3 - הכנסות ומחיקות לסירוגין
1			
2			
3			
...			
10			

2. היעזרו באקסל, איזה ביטוי אסימפטוטי תואם כל עמודה? (במונחים של  $n$ )  
 הדרכה: במקרה שמצפים לביטוי  $f(n)$ , ניתן לבדוק את מידת ההתאמה האמפירית של הנתונים על-ידי חילוק הנתונים ב  $f(n)$ . בתוכנת אקסל, למשל, ניתן לחשב קו-מגמה (trendline) ומדד ה- $R^2$  מעיד על איכות הקירוב.

### שאלה 2:

בשאלה זו נרצה לנתח את העלות של פעולות ה-join המתרחשות במהלך ביצוע concat. נגדיר את העלות של פעולת join בודדה כערך המוחלט של הפרש הגבהים של העצים שאוחדו.

- לצורך הניתוח, נבנה עצי AVL בגדלים שונים. מספר האיברים שנכניס לעץ יהיה  $n = 2^i \cdot 1000$  כאשר  $i=1, \dots, 10$ . כלומר, עבור  $i=1$  העץ בגודל 2000, ועבור  $i=10$  העץ בגודל כמיליון. את האיברים יש להכניס בסדר אקראי.
- תעדו בטבלה שלהלן את העלות הממוצעת של פעולות ה-join ואת העלות של פעולת ה-join היקרה ביותר.

מספר סידורי $i$	עלות join ממוצע עבור concat	עלות join מקסימלי עבור concat
1		
...		
10		

1. נתחו באופן תיאורטי את העלות של **join** ממוצע לשני התרחישים והסבירו אם התוצאות מתיישבות עם ניתוח הסיבוכיות התאורטי.
2. נתחו באופן תיאורטי את העלות של **join** מקסימלי באחד מהתרחישים והסבירו אם התוצאות מתיישבות עם ניתוח הסיבוכיות התאורטי.

### שאלה 3:

בניסוי זה נשווה את ביצועי עץ AVL לעומת רשימה מקושרת ומערך. ממשו רשימה באמצעות רשימה מקושרת ובאמצעות מערך (אין צורך להגיש את המימוש הזה). עבור כל  $i = 1, \dots, 10$  חזרו על התהליך הבא:

- הכניסו  $n = 1500 * i$  איברים לרשימה הממומשת בעזרת עץ AVL (לפי שלושת המקרים המפורטים בהמשך).
- הכניסו את אותם האיברים לרשימה הממומשת בעזרת רשימה מקושרת ומערך.
- עבור כל אחד מהעצים, מדדו את זמן הריצה הממוצע (הממוצע על פני  $n$  ההכנסות).

חזרו על התהליך כאשר  $i = 1500$  ההכנסות הן לפי:

1. הכנסות להתחלה – סדרת הכנסות לראש הרשימה.
2. סדרה אקראית - מכניסים במקומות אקראיים מתוך גודל הרשימה הנוכחי.
3. הכנסות לסוף – סדרת הכנסות לסוף הרשימה

רשמו את התוצאות בטבלאות הבאות:

מספר סידורי $i$	זמן ריצה בממוצע	עץ AVL הכנסות להתחלה	רשימה מקושרת הכנסות להתחלה	מערך הכנסות להתחלה
1				
2				
...				

מספר סידורי $i$	זמן ריצה בממוצע	עץ AVL הכנסות אקראיות	רשימה מקושרת הכנסות אקראיות	מערך הכנסות אקראיות
1				
2				
...				

מספר סידורי $i$	זמן ריצה בממוצע	עץ AVL הכנסות בסוף	רשימה מקושרת הכנסות בסוף	מערך הכנסות בסוף
1				
2				

			...
--	--	--	-----

מה הייתם מצפים שתהיינה התוצאות, והאם התוצאות האמיתיות מסתדרות עם ציפייה זו? הסבירו.

### הוראות הגשה

הגשת התרגיל תתבצע באופן אלקטרוני באתר הקורס במודל.  
הגשת התרגיל היא בזוגות בלבד!  
כל זוג יבחר נציג/ה ויעלה רק תחת שם המשתמש של הנציג/ה את קבצי התרגיל (תחת קובץ zip) למודל.

על ההגשה לכלול שלושה קבצים:

1. קובץ המקור (הרחבה של קובץ השלד שניתן) תחת השם AVLTreeList.py.
2. קובץ טקסט info.txt המכיל את פרטי הזוג: מספר ת"ז, שמות, ושמות משתמש.
3. מסמך תיעוד חיצוני, המכיל גם את תוצאות המדידות. את המסמך יש להגיש באחד הפורמטים הבאים: doc, docx או pdf.

שמות קובץ התיעוד וקובץ הzip צריכים לכלול את שמות המשתמש האוניברסיטאיים של הזוג המגיש לפי הפורמט ..../AVLTreeList\_username1\_username2.pdf/doc/zip בתוכן הקבצים יש לציין את שמות המשתמש, תעודות הזהות ושמות המגשים (בכותרת המסמך ובשורת הערה בקובץ המקור).

הגשת שיעורי הבית באיחור - באישור מראש בלבד. הגשה באיחור ללא אישור תגרור הורדת נקודות מהציון.  
הגשת התרגיל היא חובה לשם קבלת ציון בקורס.

בהצלחה!