# ממ"ן 14 – מבנה נתונים ומבוא לאלגוריתמים

# מטלה תכנותית – מסמך נלווה

# מגיש: עומרי ברכה 205732928

## רקע.

במטלה זו, נדרשתי בהתייחס לאלגוריתם שמקורו בממ"ן 11, לבחון מספר אסטרטגיות שונות לביצוע אותה משימה. כזכור, **מטרתו האלגוריתם הנ"ל היא למצוא את מספר האיברים השונים במערך**.

במטלה זו נבחן את הדרכים <u>השונות והמגוונות</u> לביצוע אותה משימה, תוך הסבר קצר עבור כל שיטה, ניתוח השוואות/השמות/העתקות המתבצעות בכל האלגוריתם ( אם בכלל ) וניתוח סיבוכיות זמן ריצה.

# פסאודו-קוד עבור השגרה המדוברת.

```
D(A) // A is an array of numbers
U_Size = 1
For i=2 to length(A)
U=TRUE
For j=1 to U_Size
    If A[j]=A[i]
        Then U = FALSE
        j = U_Size
    if U = TRUE

Then U_Size = U_Size +1
        A[U_Size] = A[i]
return U_Size
```

# <u>קוד המטלה.</u>

- קוד המטלה נכתב בשפת JAVA בשיטת TOP-DOWN כך שהאלגוריתמים השונים הם אלו
   הראשונים שיופיעו ואילו פונקציות העזר יהיו בתחתית העמוד.
- מצורפים בקובץ המטלה המחלקות "RBT.java", "BST.java" אלו הם הקבצים עבור המחלקות של עצים
- בממ"ן ( לדוגמא, מספר הסעיף בממ"ן ( לדוגמא, ב' אלגוריתם קיבל את השם הגנרי methodi כך ש $1 \leq i \leq 7$  הוא מספר הסעיף בממ"ן ( לדוגמא, האלגוריתם המקורי, אותו עלינו לממש בסעיף 1, קיבל את השם
  - הוגדר משתנה גלובאלי (בכדי לקבוע את private static final int MAX\_NUM = 100) בכדי לקבוע את מספר האיברים אותם נגריל באופן אקראי לבניית המערך
- עבור כל שיטה, הוגדרו גם כן המשתנים הגלובאליים אשר מטרתם לכמת את המדדים ( השוואות / השמות ), לדוגמא עבור האלגוריתם המקורי ( סעיף 1 ):

```
static int comp1 = 0;
static int ass1 = 0;
static int copy1 = 0;
```

# האלגוריתמים השונים.

# נבחן את השיטות השונות למציאת מספר האיברים השונים במערך באמצעות:

# 1. האלגוריתם המקורי –

מנקודת וובאן אנו יוצאים מנקודת int U\_size = 1, הלוא הוא הוא הוא הבשיטה זו מגדירים "סופר", הלוא שונה באופן ריק.

נרוץ על כל המערך בלולאה ובה נגדיר משתנה בוליאני כ"אמת". לאחר מכן, בתוך הלולאה הקיימת נרוץ בלולאה נוספת עד לגודל ה"סופר" שהגדרנו למעלה ונבדוק – אם איבר במקום ה-0 שווה לאיבר במקום ה-1 ( הרצה ראשונה ), נשנה את המשתנה הבוליאני לשקר ונעצור. במידה ולא כך הדבר, נצא מהלולאה הפנימית ובהכרח המשתה נשאר עדין "אמת". ובכן, מעולה, נעביר את האיברים השונים לתחילת המערך ונגדיר את ה"סופר" שלנו ב-1 וכן הלאה עד לסיום האיטרציה האחרונה. <u>הרעיון:</u> הרעיון מאחורי השגרה הוא למעשה להשוות כל איבר עם קודמו לאורך כל גודל המערך ולבדוק בעזרת משתנה עזר בוליאני, האם מתקיים מצב של שוויון בין האיברים או שמא האיברים שונים. כך אינדיקציית ה"אמת" או ה"שקר" עוזרת לנו להבין האם מהו מספר האיברים השונים במערך.

 $A \leftarrow array; |A| = n$  ניתוח זמן ריצה: בהנחה כי

כידוע, הלולאה הראשונה רצה על כל איברי המערך ( למעט הראשון ) ובמקרה הגרוע, כאשר כל איברי המערך שונים, הלולאה הפנימית תרוץ על כל איברי המערך. כאמור, מהסיבה שמדובר בלולאה מקוננת קיבלנו, במקרה הגרוע, זמן ריצה ריבועי כך –

$$0(n \cdot (n-1)) = 0(n^2 - n) \sim 0(n^2)$$

## מיון הכנסה –

<u>רקע:</u> בשיטה זו נתבסס על האלגוריתם המוכר מיון הכנסה כך שבאמצעותו נקבל מערך ממוין, נגדיר int count = 1 בלולאה על כל איברי המערך כדי לבדוק האם איבר במערך שונה מקודמו. <u>הרעיון:</u> כאשר אנו ממיינים מערך, בהכרח נוצר מצב כי כל האיברים מסודרים ב"גושים" מסוימים ועל כן מעבר על איברי המערך, אחד אחרי השני, מבטיח לנו כי לא פספסנו מקרה קצה.

לדוגמא: בהינתן המערך  $A = \{1,99,6,1\}$ , ידוע כי הפלט אמור להיות B. אם נעבור על המערך בכדי למצוא את מספר האיברים השונים מבלי למיין אותו, הפלט יהיה B ( כי הרי שבאיטרציה הלפני אחרונה, נבדוק אם B = B והתנאי יחזיר כמובן "שקר" והשגרה תעלה את ה"סופר" ב-B, אבל למעשה B כבר הופיע במערך, אבל לא הושווה ישירות אל מופעו הקודם B. לכן השימוש במיון הכנסה ( או כל מיון אחר לצורך העניין ) הכרחי להצלחת השגרה, במקרה זה.

 $A \leftarrow array$ ; |A| = n ניתוח זמן ריצה: בהנחה כי

כידוע, הלולאה בשיטה רצה על כל איברי המערך, אך אנו נעזרים בפונקציית עזר

וידוע כי זמן הריצה של מיון private static void insertionSort(int arr[]) - הכנסה הוא ריבועי ולכן בסך הכול קיבלנו

#### מיון מיזוג –

<u>רקע:</u> בשיטה זו נתבסס על האלגוריתם המוכר **מיון מיזוג** כך שבאמצעותו נקבל מערך ממוין, נגדיר int count = 1 בלולאה על כל איברי המערך כדי לבדוק האם איבר במערך שונה מקודמו. <u>הרעיון:</u> באופן זהה לחלוטין לסעיף 2, גם כאן המיון חיוני להצלחת השגרה. השוני בסעיף זה הוא כי נדרשנו למצוא <u>מיון מבוסס השוואות אופטימלי</u> ועל כן בחרתי במיון שזמן הריצה שלו אינו ריבועי, אלא קטן מכך.

$$A \leftarrow array; |A| = n$$
 ניתוח זמן ריצה: בהנחה כי

כידוע, הלולאה בשיטה רצה על כל איברי המערך, אך אנו נעזרים בפונקציית עזר

וידוע כי זמן הריצה של מיון מיזוג הוא פולינומי-לוגריתמי ולכן בסך הכול קיבלנו –

$$O(n) + O(n \cdot \log n) = O(n \cdot \log n)$$

# - של מיון מניה $\mathbf{C}$ של מיון מניה -4

<u>רקע:</u> במיון מניה, אנו למעשה נעזרים במערך "מנייה", מערך עזר בו למעשה נספרים כמות המופעים של כל איבר המערך, שהרי אם אחד מאיברי אותו מערך עזר הוא 0, אזי האיבר אליו הוא משויך אינו מופיע כלל במערך. עבור כל איבר במערך המקורי, נציין כל הופעה שלו כמספר במערך C, נעבור על איברי מערך המנייה ונבדוק האם יש איבר שאינו אפס. במידה ואין, נגדיר את ה"סופר" ובמידה ויש, כנראה שאיבר מסוים אינו מופיע במערך כלל.

<u>הרעיון:</u> כאן, בשונה מסעיפים 2-3, אין חובה למיין את המערך בכדי שהשגרה תצליח לבצע את המשימה, אפילו, אין חובה לבצע ולו השוואה אחת בין כל שני איברים. אך, כידוע אנו מוגבלים בטווח ובהיות המספרים שלמים כפי שלמדנו בספר ובמפגשים על מיון מנייה.

$$A \leftarrow array$$
;  $|A| = n$  ניתוח זמן ריצה: בהנחה כי

כידוע, הלולאה בשיטה רצה על כל איברי המערך המקורי כדי לבצע העתקה לערך מסוים במערך העזר ואז אנו רוצים על מערך העזר בלולאה נוספת כדי לבדוק. כמובן כי מיון מניה הוא מיון לינארי הבהכרח קיבלנו זמן ריצה לינארי כך –

$$0(n) + 0(n) = 0(n)$$

### **- טבלת גיבוב**

רקע: נעזר במבנה נתונים מוכר בשפת JAVA הנקרא

מבנה זה מעניק לנו פעולות מילון חיוניות למטרתנו בזמן ריצה מעולה, כגון:

HashSet<Integer> h = new HashSet<Integer>(); -עבור

# h.contains

h.add

נעבור על כל המערך בלולאה ונשתמש בפעולות המילון כדי לבדוק האם האיבר אינו נמצא בטבלת הגיבוב. במידה וכך הדבר, נגדיל את ה"סופר" ונוסיף את אותו איבר לטבלה כך שבאיטרציה הבאה, אם איבר זהה ייבדק, האלגוריתם ידע כי האיבר כבר נמצא בטבלה ועל כן, לא נגדיל את "הסופר" ולא נכניס אותו לטבלה שוב.

<u>הרעיון:</u> בעזרת מבנה נתונים מובנה של טבלת גיבוב, אנו משתמשים בבדיקה פשוטה לכל אורך המערך של איברים, העתקתם לטבלת גיבוב ע"י פעולות מילון "זולות" מבחינת זמני ריצה ובכך האלגוריתם יבצע את המשימה.

 $A \leftarrow array; |A| = n$  ניתוח זמן ריצה: בהנחה כי

(תוחלת זמן / עור: טבלת גיבוב מעניקה לו זמני ריצה טובים במקרה הממוצע

כידוע, הלולאה בשיטה רצה על כל איברי המערך המקורי כדי ומשתמשת בפעולות מילון בעלות של

- ולכן בסך הכול קיבלנו זמן ריצה לינארי המושפע בעיקר מגודל המערך 0(1)

O(n)

# **– עץ חיפוש בינארי**

<u>רקע:</u> כידוע, עץ חיפוש בינארי הוא עץ בו לכל קודקוד יש שני בנים לכל היותר, ימני ושמאלי. הבן הימני יחד עם כל צאצאיו, נמצא ביחס הסדר אחרי הקודקוד שממנו הוא יוצא ואילו הבן השמאלי וכל צאצאיו <u>קודמים</u> לקודקוד האב ביחס הסדר. גם כאן, נתבסס על פעולות המילון המהירות של העץ. במימוש המבנה, כאשר מדברים על פעולת ההוספה, העץ בודק האם האיבר שהוכנס קטן מהשורש או לא ( בין היתר ) ובכך ממשיך כדי לדעת היכן למקמו. באופן דומה, העץ בודק איפה האיבר לחיפוש נמצא. נעבור על כל המערך בלולאה ונשתמש בפעולות המילון כדי לבדוק האם האיבר אינו נמצא בעץ. במידה וכך הדבר, נגדיל את ה"סופר" ונוסיף את אותו איבר לטבלה כך שבאיטרציה הבאה, אם איבר זהה ייבדק, האלגוריתם ידע כי האיבר כבר נמצא בעץ ועל כן, לא נגדיל את "הסופר" ולא נכניס אותו לטבלה שוב

<u>הרעיון:</u> בעזרת עץ חיפוש בינארי, אנו משתמשים בבדיקה פשוטה לכל אורך המערך של איברים, העתקתם לעץ ע"י פעולות מילון "זולות" מבחינת זמני ריצה ובכך האלגוריתם יבצע את המשימה.

 $A \leftarrow array; |A| = n$  ניתוח זמן ריצה: בהנחה כי

h כידוע, זמן הריצה של כל הפעולות בעץ חיפוש בינארי תלוי בגובה העץ. נקרא לו כאשר במקרה הגרוע העץ איננו מאוזן,

:לדוגמא



ואז למעשה אין יתרון לעץ על-פני מערך לא ממוין, אך זמני הריצה עדין לינאריים. לעומת זאת, בעבור עץ בינארי מאוזן גובה העץ בהכרח  $h = \log n$  ובכך פעולות המילון שלו "יעלו" לכל היותר $\log n$ . מכאן קיבלנו שבמקרה הגרוע, זמן הריצה של האלגוריתם הינו

## **– עץ אדום-שחור** .7

<u>רקע:</u> כמה תכונות על עץ אדום-שחור

- ס הוא עץ חיפוש בינארי מאוזן, כאשר עבור כל צומת נשמור נתון נוסף הצבע שלו שיכול להיות אדום או שחור.
  - מוגדרת חוקיות לגבי אופציות הצביעה כך שנוכל להבטיח שכל מסלול מהשורש לעלה כי היה לכל היותר כפליים מכל מסלול אחר מהשורש לעלה, כך שהעץ יהיה מאוזן.
    - ס בל צומת הוא או אדום או שחור והשורש הראשי של העץ תמיד שחור 🔾
      - ס העלים הם צמתי *NULL* וצבעם תמיד שחור
        - o אם צומת הוא אדום אז שני בניו שחורים o
    - ס בל המסלולים צומת בלשהו לעלים מבילים אותו מספר של צמתים שחורים

באופן דומה לסעיף 6, נעזר בפעולות המילון אשר עץ אדום-שחור מספק לנו על מנת להוסיף איברים או לבדוק האם איברים נמצאים בעץ.

הרעיון: בעזרת עץ חיפוש בינארי, אנו משתמשים בבדיקה פשוטה לכל אורך המערך של איברים, הרעיון: בעזרת עץ חיפוש בינארי, אנו משתמשים בבחינת מני ריצה ובכך האלגוריתם יבצע את המשימה.  $A \leftarrow array; |A| = n$ ניתוח זמן ריצה: בהנחה כי

ע"פ האמור, עץ אדום-שחור הינו עץ חיפוש בינארי מאוזן ולכן במקרה הגרוע ביותר, פעולות המילון ע"פ האמור, עץ אדום-שחור הינו עץ חיפוש בינארי  $O(\log n)$  במקרה הגרוע ביותר. תיקון העץ, נעשה בזמן ריצה של - במקרה אנו משתמשים בעץ אדום-שחור נקבל סיבוכיות זמן ריצה לוגריתמית O(1)  $O(\log n)$ 

#### סיכום ביניים

#### זמני ריצה.

סיבוכיות זמן ריצה	מציאת מספר איברים שונים במערך באמצעות			
$0(n^2)$	האלגוריתם המקורי			
$0(n^2)$	מיון הכנסה			
$O(n \cdot \log n)$	מיון מיזוג			
0(n)	מערך $C$ של מיון מנייה			
0(n)	טבלת גיבוב <i>HASHSET</i>			
0(n)	עץ חיפוש בינארי			
$O(\log n)$	עץ אדום-שחור			

# נבדוק וננתח את המדדים עבור כל אלגוריתם.

מדוע המדד/ים האלו מייצגים?	מדדים ליעילות	אלגוריתם
בהתאם להסבר על האלגוריתם המקורי, כמות ההשמות של המשתנה	השוואות והשמות	האלגוריתם
boolean <b>U</b> = true;		המקורי
וההשוואות בין כל איבר במיקומים ה- $i,j$ במערך אלו המדדים המייצגים מהסיבה		
שבמקרה הגרוע, כל אחד מהם ישפיע באופן ישיר על זמן הריצה של המערך		
כידוע, מיו הכנסה הינו מיון מבוסס השוואות ובמיון זה מתבצעת השמה	השוואות והשמות	מיון הכנסה
arr [i+1] = arr [i];		
ולכן יעילותו אלגוריתם ווא אלגוריתם וולכן יעילותו אלגוריתם ווואלגוריתם ווואלגוריתם ווואלגוריתם ווואלגוריתם ווואל		
מהתהליכים המתרחשים במיון הכנסה		
שימוש במיון מיזוג, בו אנו נעזרים במערך עזר, משווים איברים, מבצעים השמות	השוואות, השמות	מיון מיזוג
ומעתיקים למערך העזר ערבים, למעשה משפיע על יעילותו של האלגוריתם ה-	והעתקות	
באופן ישיר <u>net.hod3</u>		
כפי שלמדנו, מיון מניה הינו מיון בזמן <u>לינארי שאינו מבוסס השוואות</u> . אנו נעזרים	העתקות	מערך $\emph{C}$ של מיון
במערך עזר, אשר הוא מייצג את כמות המופעים של כל איבר במערך המקורי.		מנייה
לכן, הפעולה היחידה שנעשית היא למעשה <u>העתקה</u>		
העתקות בשימוש בטבלת גיבוב, אנו למעשה מעתיקים את איברי המערך לטבלה כל עוד		טבלת גיבוב
טרם הופיעו ובכך המדדים המייצגים בשיטה זו הם השוואות והעתקות		HASHSET
באופן זהה לטבלת גיבוב, כך גם בעץ חיפוש בינארי ובעץ אדום-שחור. ההבדל	השוואות והעתקות	עץ חיפוש בינארי
היחיד הוא שעל עץ חיפוש בינארי ועץ אדום-שחור מתבצעות באופן ניכר יותר	השוואות והעתקות	עץ אדום-שחור
השוואות מהסיבה שהכנסת איבר לעץ, מלווה בלא מעט הגבלות ( בשל המבנה		
הייחודי של כל עץ )		

#### N=20,100,1000,10000,100000ע"פ המטלה נבחן דוגמאות הרצה עבור

For N = 100: The Original Algorithm performs: The Original Algorithm performs: Compares: 47611 Assignments: 1098 Copies: 0 Total distinct numbers: 100 Compares: 500717 Assignments: 10098 Compares: 3003 Assignments: 162 Copies: 0 Total distinct numbers: 100 Copies: 0 Total distinct numbers: 64 The 2nd Algorithm performs: Compares: 9999 Assignments: 25023946 Copies: 0 Total distinct numbers: 100 Compares: 999 Assignments: 249991 Compares: 99 Assignments: 2715 Copies: 0 Fotal distinct numbers: 100 Copies: 0 Total distinct numbers: 64 The 3rd Algorithm performs: Compares: 88081 Assignments: 133616 Copies: 133616 Total distinct numbers: 100 The 3rd Algorithm performs: Compares: 442 Assignments: 672 Copies: 672 Total distinct numbers: 64 The 3rd Algorithm performs: Compares: 6640 Assignments: 9976 Copies: 9976 Total distinct numbers: 100 The 4th Algorithm performs: The 4th Algorithm performs: Compares: 0 Assignments: 0 Copies: 1000 Total distinct numbers: 100 Assignments: 0 Copies: 100 Total distinct numbers: 64 Assignments: 0 Copies: 10000 Fotal distinct numbers: 100 The 5th Algorithm performs: Compares: 10000 Assignments: 0 The 5th Algorithm performs: Compares: 100 Assignments: 0 Compares: 1000 Assignments: 0 Copies: 100 Fotal distinct numbers: 100 Copies: 64 Total distinct numbers: 64 Copies: 100 Total distinct numbers: 100 The 6th Algorithm performs: Compares: 103320 Assignments: 0 Copies: 100 Total distinct numbers: 100 The 6th Algorithm performs: Compares: 8220 Assignments: 0 Copies: 64 Total distinct numbers: 64 he 6th Algorithm performs: Compares: 1006952 Assignments: 0 Copies: 100 Total distinct numbers: 100 The 7th Algorithm performs: The 7th Algorithm performs: Compares: 1006952 Assignments: 0 Copies: 100 Total distinct numbers: 100 The 7th Algorithm performs: Compares: 8220 Assignments: 0 Compares: 103320 Assignments: 0 Copies: 100 Total distinct numbers: 100 Copies: 64 Total distinct numbers: 64

For N = 100000: For N = 1000000: The Original Algorithm performs: The Original Algorithm performs: Compares: 5046117 Assignments: 100098 Compares: 50471214 Copies: 0 Total distinct numbers: 100 Assignments: 1000098 Copies: 0 Total distinct numbers: 100 The 2nd Algorithm performs: Compares: 99999 Assignments: 2472607852 The 3rd Algorithm performs: Copies: 0 Total distinct numbers: 100 Compares: 19620184 Assignments: 19951424 The 3rd Algorithm performs: Copies: 19951424 ompares: 1047339 Total distinct numbers: 100 Assignments: 1668928 Copies: 1668928 Total distinct numbers: 100 The 4th Algorithm performs: Compares: 0 The 4th Algorithm performs: Assignments: 0 Compares: 0 Copies: 1000000 Assignments: 0 Total distinct numbers: 100 Total distinct numbers: 100 The 5th Algorithm performs: The 5th Algorithm performs: Compares: 100000 Compares: 1000000 Assignments: 0 Copies: 100 Assignments: 0 Copies: 100 Total distinct numbers: 100 Total distinct numbers: 100 Compares: 9997764 Assignments: 0 The 6th Algorithm performs: Copies: 100 Compares: 13059381 Total distinct numbers: 100 Assignments: 0 The 7th Algorithm performs: Copies: 100 Compares: 9997764 Assignments: 0 Total distinct numbers: 100 Copies: 100 Total distinct numbers: 100

#### דוגמאות הרצה נוספות –

```
he array is:
50 27 49 55 91 24 33 58 84 33 52 23 67 52 58 28 24 92 58 50
                                                              The Original Algorithm performs:
The Original Algorithm performs:
                                                               ssignments: 32
                                                               opies: 0
otal distinct numbers: 14
Copies: 0
Total distinct numbers: 4
Compares: 4
Assignments: 15
Copies: 0
Total distinct numbers: 4
                                                              Compares: 19
Assignments: 120
Copies: 0
Total distinct numbers: 14
                                                              The 3rd Algorithm performs:
The 3rd Algorithm performs:
                                                               Compares: 62
Assignments: 88
Copies: 88
Total distinct numbers: 14
                                                              The 4th Algorithm performs:
Copies: 5
Total distinct numbers: 4
                                                               opies: 20
otal distinct numbers: 14
The 5th Algorithm performs:
                                                              The 5th Algorithm performs:
Compares: 20
Assignments: 0
Copies: 14
                                                               otal distinct numbers: 14
The 6th Algorithm performs:
                                                              The 6th Algorithm performs:
Compares: 361
Compares: 19
Assignments: 0
Copies: 4
Total distinct numbers: 4
                                                              ompares: 361
Assignments: 0
Opies: 14
Total distinct numbers: 14
The 7th Algorithm performs:
Compares: 361
Assignments: 0
Copies: 14
  otal distinct numbers: 4
```

# <u>מסקנות.</u>

כפי שניתן להסיק, ככל שגדל המערך, באופן די טריביאלי כל שיטה "תעלה" יותר זמן, אך באופן די מדהים, לאחר שבחנו 7 שיטות שונות, אנו רואים כי כל שיטה "תעלה" באופן <u>שונה.</u> מכאן אנו רואים כי למדדים המייצגים יש השפעה מכרעת בזמני הריצה של האלגוריתמים למציאת מספר המספרים השונים במערך. לדוגמא, כאשר נשתמש במיון לינארי ( שאינו מבוסס השוואות ), כאשר אנו מעתיקים בלבד את ערכי המערך למערך עזר ומגלים כך את מספר המופעים, קל לראות כי אכן מדובר בזמן ריצה לינארי למספר המופעים וכי מספר ההעתקות גדל בהתאם לגודל המערך ( באופן ישיר ).

### <u>סיכום.</u>

לסיכום, בחנתי 7 שיטות שונות למציאת המספרים השונים במערך. חלקן מהירות יותר וחלקן מהירות פחות. תיארתי בקצרה את נכונות האלגוריתמים, את זמני הריצה ואת הרעיון מאחוריהם. הראיתי דוגמאות הרצה כנדרש ואף צירפתי סרטון המחשה שהכנתי על "מאחורי הקלעים" של האלגוריתם method4 , אשר משתמש במערך C של מיון מניה.

#### מקרא קובץ.

- 1. קובץ המטלה "*Matal14*"
- 2. קבצי מחלקות העזר ( עץ חיפוש בינארי ועץ אדום שחור )
  - 3. סרטון הממחיש את פעולת האלגוריתם method4