תרגיל בית 2 – מבני נתונים

מגישים : יהונתן לייטנר 312474646 , בראל לנציאנו 206082794 , נתנאל שאשא 313298473

שאלה 1

ב.2 – סיבוכיות הקוד שרשמתי הינו **ON** - הקוד בנוי מ-2 לולאות , האחת שעוברת על הקוד ומכניסה למחסנית את הסוגרים הפותחים , והשנייה (שלא עובדת בתוכה) עוברת על הסוגרים הסוגרים ומשווה אותם לאיבר העליון במחסנית שעובד בזמן ריצה קבוע.

ב.3 – שימוש במחסנית מאפשר מימוש יעיל של בעיה זו מכיוון שסדר הסוגרים הנכון מוכנס בצד אחד של המחסנית , ובאמצעות כך מאפשר השוואה בזמן ריצה קבוע של איברים מהסטרינג עצמו אל המחסנית.

במידה ולא ההינו משתמשים במחסנית פעולת ההשוואה הייתה מצריכה ביצוע פעולות השוואה נוספות , דבר אשר היה מעלה את סיבוכיות הקוד.

שאלה 2

3. האלגוריתם מקבל מטריצה A ומספר מזהה של עיר. מאתחלים 3 משתנים : הראשון – מספר הערים במטריצה , השני – מספר העיר המקורי שקיבלנו כחלק מהקלט והשלישי המרחק הכולל שעברנו. בנוסף הגדרנו ערימת מינימום ריקה , שתשמש לברירה בין אופציות ההליכה , רשימה ריקה לערים בהם נעבור ורשימה זמנית לטובת חישוב.

האלגוריתם מתחיל לרוץ בלולאת WHILE כל עוד מספר הערים שונה מ0.

נתחיל לולאת FOR פנימית , כאשר בכל לולאה נכניס את הנתונים שמתאימים לעיר המקורית לתוך הערימה , תוך תנאי שאנחנו לא מכניסים לערימה ערים שכבר היינו בהם (באמצעות הכנסת הערים לרשימה ותנאי על הלולאה). כאשר הלולאה מסתיימת ניקח את הערך הנמוך ביותר בערימה ונוסיף אותו למרחק הכולל שעברנו ונוריד את מספר הערים באחד. כך נצא לבסוף מהלולאה ונקבל את המבוקש – נחזיר את מזהה העיר שבו היינו ואת המרחק הכולל שעברנו.

4. סיבוכיות הקוד שכתבנו הינו **NLOGN** מכיוון שכל פעם אנחנו לוקחים חלק מסוים והקלט ומבצעים עליו פעולות N פעמים.

שאלה 3

סעיף ה:

מדד היעילות מסייע לנו לקבוע את יעילות הפונקציה. ככל שהמדד נמוך יותר, הפונקציה יעילה יותר. (מדד 1.0 הוא המדד הטוב ביותר.)
לכל HASH Table מספר פעולות שונה.

	סט נתונים	פונקציית גיבוב	התמודדות עם התנגשויות	מדד היעילות
1	X	חלוקה	שרשור	1.0672268907563025
2		חלוקה	בדיקה ריבועית	2.042016806722689
3		חלוקה	גיבוב כפול	2.0588235294117645
4		הכפלה	שרשור	1.0756302521008403
5		הכפלה	בדיקה ריבועית	2.319327731092437
6		הכפלה	גיבוב כפול	2.0840336134453783
1	Y	חלוקה	שרשור	1.0
2		חלוקה	בדיקה ריבועית	1.0
3		חלוקה	גיבוב כפול	1.0
4		הכפלה	שרשור	1.0
5		הכפלה	בדיקה ריבועית	3.2941176470588234
6		הכפלה	גיבוב כפול	1.7899159663865547
1	Z	חלוקה	שרשור	9.470588235294118
2		חלוקה	בדיקה ריבועית	14.546218487394958
3		חלוקה	גיבוב כפול	3.53781512605042
4		הכפלה	שרשור	1.0168067226890756
5		הכפלה	בדיקה ריבועית	1.7647058823529411
6		הכפלה	גיבוב כפול	1.7563025210084033

<u>נתונים X :</u>

בקובץ נתונים X ישנם מספרים בעלי 9 ספרות הרשומים באופן אקראי ובלתי תלוי, לכן היעילות היא יחסית גבוהה בכל אחת משיטות הגיבוב.
כאשר התמודדנו עם ההתנגשויות בעזרת שרשור ,מדד היעילות היה דומה ונמוך ביחס לטבלאות האחרות. ככל הנראה זה נגרם עקב חלוקה אקראית של המספרים . לעומת זאת , בפונקציות האחרות המדד היה גבוה מ-2 . כלומר, נוצרו התנגשויות שאילצו את הפונקציות למצוא מקום אחר. והובילו לחוסר יעילות.

<u>נתונים Y:</u>

מכיל מספרים ,בסדר עולה מ-1 עד 119 (אורך סט הנתונים). מכיוון ש-m=149 לא נוצרות בכלל התנגשויות ולכן בשיטת **החלוקה**, טבלאות 1, 2 ו-3, אין צורך לטפל בהן ומדד היעילות עומד על 1.0. בנוסף גודל סט הנתונים (119) קטן מ-m (149) ולכן הטבלה לא תתמלא- שארית החלוקה הינה המספר עצמו.

לעומת זאת,
כשאר השתמשנו בשיטת ההכפלה והתמודדנו עם ההתנגשויות בשיטת הבדיקה הריבועית ושיטת גיבוב כפול , התוצאות היו פחות טובות- גבוהות מ-1

<u>ככל הנראה בגלל שמודבר במספרים עוקבים השימוש בערך תחתון בנוסחה הוביל להחזרת תוצאות דומות, ולכן להתנגשויות רבות.</u>

<u>נתונים Z :</u>

בקובץ זה ישנם מספרים בקבוצות של 6 מספרים עוקבים (מלבד הקבוצה האחרונה שמכילה רק 5 מספרים עוקבים).
בשימוש בשיטת החלוקה כל קבוצת מספרים שכזו מתחלקת לשישה מספרים בין 1 ל-6, כלומר כאשר מבצעים h(x)%m ו-m קבוע בכל הטבלאות ושווה ל-149 (מספר ראשוני).
התוצאות הן זהות לכל קבוצה דבר היוצר התנגשויות רבות (כ-20 לכל מספר) ולכן תוצאות היעילות הנגזרות ממספר הפעולות לסט נתונים זה בשיטת החלוקה גבוהות ביחס לשיטת ההכפלה. בשיטת ההכפלה יש פיזור טוב יותר ולכן גם מדד היעילות נמוך יותר (כלומר טוב יותר).
ולכן במקרה זה נעדיף לעבוד עם שיטת ההכפלה.

<u>סעיף ו'</u>

כאשר משתמשים בשיטת החלוקה וההכפלה , נשים לב כי הגדלת ערך הפרמטר M מצמצם את מספר ההתנגשויות.
במקרה שלנו מספר האברים במילון קבוע ושווה ל-119 (גודל המילון).
ולכן ככל שנגדיל את M נשפר את היעילות של הקוד.

עבור הנתונים X :
במקרה זה בחרנו להגדל את M מ-149 ל233 כאשר השתמשנו בשיטת גיבוב של חלוקה והתמודדנו עם שגיאות בעזרת שרשור.
בחרנו ב233 כיוון שהוא מספר ראשוני הגדול מ149 , ולכן בעזרתו נוצרות פחות התנגשויות והיעלות גדלה ל-1.025210084 (לעומת 1.06722689 עבור M=149)

עבור הנתונים Y :
עבור הנתונים ב-Y קיבלנו מדד יעלות של 1.0 עבור כל סוגי הגיבוב וההתמודדות עם ההתנגשויות. זהו המדד הגבוה ביותר ולכן אין אנו יכולים לשפר אותו באמצעות הגדלת ארך ה-M .

עבור הנתונים Z:
עבור נתונים אילו בחרנו לשפר את מדד היעילות באמצעות שינוי הערך של A.
לאחר בדיקה במקורות שונים באינטרנט, גילנו כי במקרה זה כדי לשנות את -A לערך 0.6180339887 . ואכן במימוש גיבוב הכפלה והתמודדות עם התנגשויות בשרשור מדד היעילות השתפר ל 1.0 (המדד הטוב ביותר)