מבני נתונים - תרגיל מעשי 2

סמסטר ב' תש"ף

מטרת תרגיל זה הינה להבין לעומק את דרך הפעולה של טבלאות hash עם פתרון התנגשויות בשיטת open addressing, ולהשוות בין הגרסאות השונות של שיטה זו כפי שנלמדו בשיעור.
בתרגיל זה תממשו טבלאות hash עם open addressing בשיטות הבדיקה (probing) הבאות: linear probing, quadratic probing, double hashing.

חלק א – מימוש

בחלק זה בתרגיל תממשו טבלת hash. כניסה של איבר בטבלת ה-hash היא מהטיפוס HashTableElement, המכיל שדה key (המפתח של האיבר) ו-value (הערך הנוסף). הממשק שאותו הטבלה צריכה לממש הוא IHashTable, המכיל את המתודות הבאות:

- Insert(HashTableElement hte) – מכניסה את האיבר hte לטבלה אם קיים מקום פנוי בסדרת החיפוש של hte.key ולא קיים בטבלה איבר בעל אותו המפתח. אם קיים בטבלה איבר בעל אותו המפתח, יזרק חריג מסוג KeyAlreadyExistsException. אחרת, אם לא קיים מקום פנוי בסדרת החיפוש, הפונקציה תזרוק TableIsFullException.
- Find(long key) – מחזירה את ה-HashTableElement עם המפתח key, אם קיים. מחזירה null אחרת.
- Delete(long key) – מוחקת את האיבר עם המפתח key מהטבלה, אם קיים. אחרת, זורקת KeyDoesntExistException.

1. <u>מימוש טבלת Hash עם Open Addressing</u>.

ממשו את המחלקה האבסטרקטית OAHashTable. מחלקה זו מממשת את הממשק IHashTable, וקיימת לה מתודה אבסטרקטית יחידה שהיא Hash(long key, long i). מתודה זו תממש, עבור מחלקות הבת של OAHashTable, את מציאת האיבר ה-i בסדרת החיפוש של key – כלומר, היא מחזירה אינדקס לתוך טבלת החיפוש.

הבהרה: סדרת החיפוש של מפתח k בשיטת open addressing היא סדרת האינדקסים

$$h(k,0), h(k,1), h(k,2), \dots, h(k,m-1)$$

עליכם אם כן לממש את כל הפונקציות של טבלת החיפוש בהסתמך על הפונקציה האבסטרקטית hash.

בנוסף, OAHashTable מגדירה קונסטרקטור שמקבל את גודל המערך m כפרמטר (זכרו שיש משמעות להגדרת קונסטרקטור למחלקה אבסטרקטית – קונסטרקטור זה יקרא מתוך מחלקות הבת הקונקרטיות).

2. <u>מימוש טבלאות Hash עם מגוון Probing Schemes</u>.

בשאלה זו עליכם לממש מספר מחלקות קונקרטיות היורשות מ-OAHashTable. כל המחלקות הללו ישתמשו במשפחה האוניברסלית של פונקציות לינאריות מהצורה $ax + b$ מודולו ראשוני p ומודולו גודל הטבלה m, כפי שנלמדו בכיתה.

א. ממשו את המחלקה ModHash, המייצגת פונקציית hash מהמשפחה. למחלקה יהיו את המתודות הבאות:
   - המתודה Hash(long x), שתשערך את הפונקציה על המפתח x.
   - המתודה הסטטית GetFunc(int m, long p), שתחזיר אוביקט ModHash המייצג פונקציה שנבחרה אקראית מתוך המשפחה.

ב. ממשו את המחלקות הקונקרטיות הבאות היורשות מ-OAHashTable. המחלקות ישתמשו בפונקציות hash מתוך המחלקה ModHash. הקונסטרטור של כל מחלקה יקבל את גודל הטבלה m ואת הראשוני p כאשר $p \geq m$. המחלקות הינן:
   - LPHashTable – טבלה עם linear probing.
   - QPHashTable – טבלה עם quadratic probing. הפונקציה שבה נשתמש עבור קביעת האינדקסים היא
     $$h(k,i) = (h'(k) + i^2) \bmod m$$
     כאשר h' היא פונקציית hash מתוך המשפחה האוניברסלית הנ"ל.
   - AQPHashTable – טבלה עם alternating quadratic probing. כלומר, הפונקציה שבה נשתמש עבור קביעת האינדקסים היא
     $$h(k,i) = (h'(k) + (-1)^i \cdot i^2) \bmod m$$
     כאשר h' היא פונקציית hash מתוך המשפחה האוניברסלית הנ"ל.
   - DoubleHashTable – טבלה עם double hashing. כלומר, הפונקציה שבה נשתמש עבור קביעת האינדקסים היא
     $$h(k,i) = \left(h_1'(k) + i \cdot h_2'(k)\right) \bmod m$$
     כאשר h_1', h_2' הן פונקציות hash הנבחרות מתוך המשפחה האוניברסלית הנ"ל (באופן בלתי תלוי).

חלק ב - ניסויים

3. בשאלה זו נשווה בין quadratic probing ל-alternating quadratic probing, ונבין את הקשר בין בחירת המקדמים של הביטוי הריבועי לבין התקינות של סדרת הבדיקה.

א. עבור המספר הראשוני $q = 6571$, חשבו אמפירית את גדלי הקבוצות:
$$Q_1 = \{i^2 \bmod q \mid 0 \leq i < q\}$$
$$Q_2 = \left\{(-1)^i \cdot i^2 \bmod q \,\middle|\, 0 \leq i < q\right\}$$

ב. חזרו על השלבים הבאים 100 פעמים:
   i. צרו טבלה חדשה מהמחלקה QPHashTable כאשר $m = 6571$ ו-$p =$ 1,000,000,007.
   ii. הכניסו את m איברי הסדרה הרנדומית $(a_i)_{i=0}^{m-1}$ לתוך הטבלה, כאשר
      1. $a_i = 100i + b_i$, ו-b_i מתפלג אחיד בטווח [0,99] עבור כל $0 \leq i \leq m - 1$.

האם כל פעולות ההכנסה הושלמו בהצלחה, או שנזרקו חריגים? חזרו על התהליך הקודם עם AQPHashTable במקום QPHashTable. כיצד ניתן להסביר את השוני בין התוצאות?

ג. (בונוס) למדו על שאריות ריבועיות (quadratic residues) והסבירו את התופעה שבתרגיל זה. האם היא הייתה מתרחשת לכל ראשוני שהיינו בוחרים? מהו התנאי לקיום התופעה?

4. בשאלה זו נשווה בין המימושים השונים ל-open addressing. בצעו את המדידות הבאות עבור כל אחד מסוגי הטבלה LPHashTable, QPHashTable, AQPHashTable, DoubleHashTable. תעדו את זמן הריצה של כל סעיף בטבלה של אותו הסעיף. עבור כל אחד מהסעיפים, הוסיפו הסבר מילולי להבדלים בזמני הריצה בין סוגי הטבלאות. <u>בשאלה זו לא אמורים להיזרק חריגים</u>.

א. צרו טבלה מגודל $m = 10{,}000{,}019$ כאשר $p = 1{,}000{,}000{,}007$. הכניסו לטבלה את $n = \left\lfloor \frac{m}{2} \right\rfloor$ איברי הסדרה הרנדומית $(a_i)_{i=0}^{n-1}$, כאשר $a_i = 100i + b_i$, ו-b_i מתפלג אחיד בטווח $[0,99]$ עבור כל $0 \leq i \leq n-1$.

Class	*Running Time*
LPHashTable	
QPHashTable	
AQPHashTable	
DoubleHashTable	

ב. חזרו על הסעיף הקודם, אבל כש-$n = \left\lfloor \frac{19m}{20} \right\rfloor$. אין לבצע סעיף זה עבור QPHashTable (נמקו מדוע). האם ההבדל בביצועים לעומת הסעיף הקודם שונה בהתאם לסוג הטבלה? נמקו.

Class	*Running Time*
LPHashTable	
AQPHashTable	
DoubleHashTable	

5. בשאלה זו נחקור את השפעת מחיקת איברים ב-open addressing על סיבוכיות הזמן של פעולות על הטבלה.צרו איבר של המחלקה DoubleHashTable עבור $m = 10{,}000{,}019$ ו-$p = 1{,}000{,}000{,}007$, ובצעו את התהליך הבא 6 פעמים:

a. הגרילו את הסדרה הרנדומית $(a_i)_{i=0}^{n-1}$ כבשאלה הקודמת, כאשר $n = \left\lfloor \frac{m}{2} \right\rfloor$.
b. הכניסו את איברי הסדרה לטבלה.
c. מחקו את איברי הסדרה מהטבלה.

שימו לב שהסדרה הרנדומית מוגרלת מחדש בכל איטרציה. השוו את זמן ביצוע 3 האיטרציות הראשונות לזמן ביצוע 3 האיטרציות האחרונות. האם קיים הבדל? אם כן, הסבירו מדוע.

Iterations	*Running Time*
First 3 iterations	
Last 3 iterations	

הגשה

הגשת התרגיל תתבצע באופן אלקטרוני באתר הקורס במודל.

הגשת התרגיל היא בזוגות בלבד!

הגשה ביחידים תתאפשר רק באישור המתרגלים.

כל זוג ייבחר נציג **אחד** ויעלה תחת שם המשתמש שלו את קבצי התרגיל (תחת קובץ zip) למודל. על ההגשה לכלול <u>10</u> קבצים:

1. שמונת קבצי המקור שניתנו תחת שמותיהם המקוריים.
2. קובץ טקסט info.txt המכיל את פרטי המגישים הבאים: תז, שמות ושמות משתמש.
3. מסמך תיעוד חיצוני, המכיל גם את תוצאות המדידות. את המסמך יש להגיש באחד הפורמטים הבאים: txt, rtf, doc, docx או pdf.

שמות קובץ התיעוד וקובץ הzip צריכים לכלול את שמות המשתמש האוניברסיטאיים של **שני המגישים** לפי הפורמט ...HashTable_username1_username2.pdf/doc/zip/. בתוכן הקבצים יש לציין את שמות המשתמש, תעודות הזהות ושמות המגישים (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!