

מבני נתונים - פרויקט מס' 2 – ערכות מתקדמות

Amit Kacen | 212677884 | amitkacen

Dan Remeniuk | 322377490 | danremeniuk

חלק מעשי – תיעוד

בפרויקט זה כתבנו מחלקה בשם Heap המממשת סוג עירומות שונות מעל מספרים שלמים חיוביים. המחלקה תומכת בפעולות מימוש ואפשרות לבחור (בעת יצירת העירמה) האם להשתמש ב lazy melds האם להשתמש ב lazy. מטרת המחלקה היא לאחסן זוגות של Key-Value ולתמוך בעפעולות עירמה ייעילות.

תפקידו של כל חבר במחלקה

- self.lazyMelds – שדה בוליאני הקובע האם הירימה עבדת במצב של "מייזג עצל". אם הערך true, פעולות מייזג והכנסה הן מהירות והסידור נדחה. אם false, מתבצע סידור (successive linking) בכל פעולה הכנסה ומייזג.
- self.lazyDecreaseKeys – שדה בוליאני הקובע את אופן הטיפול בהקטנת מפתחה. אם הערך true, מתבצע שימוש במנגנון Cascading Cuts של עירמת פיבונאצ'י. אם false, מתבצעת פעולה Heapify-Up סטנדרטית.
- self.min – שדה המחזיק מצביע לאובייקט בעל המפתח המינימלי הנוכחי בעירמה.
- self.head – שדה המחזיק מצביע לאייר הראשון ברשימת השורשים של הירימה.
- self.last – שדה המחזיק מצביע לאייר האחרון ברשימת השורשים של הירימה.
- self.size – שדה המחזיק את מספר האיברים הכלול בעירמה.
- self.numTrees – שדה המחזיק את מספר העצים הנוכחיים ברשימת השורשים.
- self.numMarkedNodes – שדה המונונה את כמות הצלמתים המסומנים בעירמה.
- self.totalLinks – מוניה הסופר את סך פעולות הקישור (Link) שבוצעו בין שני עצים לאורך ח' הירימה.
- self.totalCuts – מוניה הסופר את סך פעולות החיתוך (Cut) שבוצעו לאורך ח' הירימה.
- self.totalHeapifyCosts – מוניה הסופר את סך פעולות ההחלפה (Swap) שבוצעו במסגרת תהליך Heapify-Up.

תיאור הפעולות

- **insert** – הפעקציה מכניסה אייר חדש לעירמה עם המפתח והמידע הנתונים. היא מוחזירה את האובייקט HeapItem החדש שנוצר, המכיל מצביע לצומת בעירמה. מבחינת אופן הפעולה, הפעקציה יוצרת תחילת צומת חדש ומתחילה את מצביעי next וה-prev שלו להצביע על עצמו. לאחר מכן, היא יוצרת עירמה זמנית חדשה (heap2) המכילה רק את האיבר החדש, וביצעת קריאה לפונקציה meld כדי למציג את הירימה הזמנית לתוך הירימה הנוכחיית.
ניתוח סיבוכיות: מן הריצה תלו依 בערך של השדה lazyMelds:
 - אם lazyMelds הוא true, פעולה meld-המבצע שרשור של רשימות השורשים בלבד, ללא succesiveLinking. פעולה זו כוללת מספר קבוע של עדכוני מצביעים. לכן, זמן הריצה הוא $O(1)$.
 - אם lazyMelds הוא false, פעולה meld-המבצע קוראת ל-successiveLinking. במקרה הגראן ביותר (בו חיבור העץ החדש גורר רצף של מייזגים), יתכונו מייזגים עד לגובה העץ המקסימלי. מאחר וגובה העץ חסום לוגריתמית, זמן הריצה במקרה הגראן הוא $O(logn)$.

- findMin** - הfonקציה מחזירה את האובייקט בעל המפתח המינימלי בערימה. הfonקציה ניגשת `self.min`, אשר מחזיק מצביע לאיבר המינימלי הנוכחי. אם הערימה ריקה, מוחזר ערך `null`.

ניתוח סיבוכיות: המצביע למינימום מתחזק באופן שוטף ולכן הגישה אליו קוראת בזמן קבוע של $O(1)$.
- deleteMin** - הfonקציה מוחקת את האיבר המינימלי מהערימה. תחילת הfonקציה מטפלת במקרים קצה (ערימה ריקה או ערימה עם איבר יחיד). לאחר מכן, היא מוציאה את איבר המינימום משימוש השורשים וمعدכנת את המצביעים ואת גודל הערימה בהתאם. אם לאיבר שנמוך היו ילדים, הfonקציה הופכת אותם לשורשים עצמאיים. לבסוף, הfonקציה קוראת `successiveLinking` כדי לבצע פועלות `consolidate` ולסדר את הערימה מחדש, ואז סורקת את רשימת השורשים החדשה כדי למצוא ולעדכן את מצביע `min` החדש.

ניתוח סיבוכיות: זמן הריצה תלוי בכמות העצים ברשימה השורשים ובדרגת האיבר המינימלי שנמוך.

 - במקרה של `lazyMelds=true`, יתכן שרישימת השורשים מכילה (n) עצים (עקב רצף הכנסות ללא סידור). פועלות `successiveLinking` ועדכון `min` יאלצו לסרוק את כל העצים הללו. لكن, זמן הריצה במקרה הגרוע הוא (n) . עם זאת, ניתן כי עלות `amortize` במצב זה היא $(log n)$.
 - במקרה של `lazyMelds=false`, רשימת השורשים תמיד מכילה לכל היותר $(log n)$ עצים. لكن, תהילך המיחקה, הקישור מחדש ועדכון המינימום חסומים כולם על ידי גובה העץ. זמן הריצה הוא $(log n)$ בWC.
- decreaseKey** - הfonקציה מקטינה את המפתח של איבר נתון (x) בערך מסוים ($diff$) ושומרת על חוקיות הערימה. תחילת הfonקציה מעדכנת את המפתח של האיבר ובודקת אם הוא קטן מהמינימום הנוכחי (ומעדכנת את `min` בהתאם). לאחר מכן, באופן הופורה תליי בדגל `cascadingCut`:

 - אם הוא `true`, נבדק האם הופר כל הערימה. במידה וכן, נקראת הfonקציה `cascadingCut` כדי לנתק את הצומת ולטפל בשרשראת הסימונים במעלה העץ.
 - אם הוא `false`, נקראת הfonקציה `heapifyUp` כדי לבצע עפוע של האיבר לפני מעלה.

ניתוח סיבוכיות: נראה כי זמן הריצה בWC הוא $(log n)$

 - במקרה הראשון (if): הfonקציה מבצעת `cascadingCut`. במקרה הגרוע, בהתאם לערך של `lazyMelds`, עלות הריצה היא $(n^2 \log n)$ או $(log n)$.
 - במקרה השני (else): הfonקציה מבצעת `heapifyUp` שעולותה חסומה בגובה העץ ולכן זמן הריצה יהיה $(log n)$.
- heapifyUp** - פונקציית עזר המעליה צומת במעלה העץ עד שחוקיות הערימה משוחזרת. הfonקציה רצתה בלולאה כל עוד הצומת אינו שורש ומפתחו קטן מהמפתח אבי. בכל איטרציה מתבצעת החלפה (באמצעות קרייה `swapWithParent`) והצומת מתקדם לאביו.

ניתוח סיבוכיות: במקרה הגרוע, הצומת יעליה מהעליה העמוק ביותר ועד לשורש. על כן, זמן הריצה בWC הוא $(log n)$.
- swapWithParent** - פונקציית עזר המבצעת החלפה בין צומת לאביו. ההחלפה מתבצעת על ידי החלפת התוכן של `HeapItem` בין שני הcents, ועדכון המצביעים המתאימים בתוך ה-`item`. כמו כן, הfonקציה מעדכנת את המצביעים הגלובליים (`head, last, min`) במידת הצורך.

ניתוח סיבוכיות: הfonקציה מבצעת פעולות בזמן קבוע בלבד, ולכן זמן הריצה הוא (1) .
- cascadingCut** - פונקציה רקורסיבית המבצעת את מגנון הניתוק המדורה. תחילת הfonקציה קוראת `cut` כדי לנתק את הבן (x) מאביו (y). לאחר מכן היא בודקת את האב: אם האב אינו שורש והוא אינו מסומן (`marked`), נסמן אותו `marked=true` והתהילך נעצר. אם האב כבר מסומן, הסימון מוסר (`marked=false`), והfonקציה קוראת לעצמה רקורסיבית כדי לנתק גם את האב מאביו שלו.

ניתוח סיבוכיות: במקרה הגרוע ביותר, שרשרת הניתוקים נמשכת עד לשורש העץ. מכיוון שגובה העץ חסום לוגריתמי, מספר הקרייאות הרקורסיביות חסום על ידי גובה העץ. כאשר `marked=True`, זמן הריצה הוא $(log n)$, אחרת כל פעולה `cut` היא בזמן ריצה של $(log n)$ ולכן סה"כ $(n^2 \log n)$.

- cut** - פונקציית עזר המנטקת קשר בין צומת (x) לאביו (y) והופכת את x לשורש חדש בעירימה.

הfonקציה מעדכנת את מצביעי הילדים של האב, מקטינה את דרגתו, ומאפשרת את סימון ה-*marked* של הבן. לאחר מכן, הבן מתווסף לרשימה השורשים של העירימה באמצעות יצירת עירימה חדשה וקירה ל-*meld*.

ניתוח סיבוכיות: הfonקציה מבצעת מספר קבוע של עדכוני מצביעים וקוראת ל-*meld*. בהנחה שהעירימה מוגדרת עצלה (*lazyMelds=true*), פעולה המיזוג היא $O(1)$ וכן גם זמן הריצה של המתודה. במידה I , *lazyMelds=false*, הסיבוכיות תהיה $O(log I)$.
- delete** - הfonקציה מוחקת איבר כללי (x) מהעירימה. השימוש מתבצע על ידי הקטנת המפתח של האיבר לערך המינימלי האפשרי כדי להבטיח שהוא יהיה הקטן ביותר בעירימה. לאחר שינוי המפתח, הfonקציה קוראת ל-*deleteMin* כדי להוציא את האיבר מהעירימה.

ניתוח סיבוכיות: הfonקציה מורכבת משני שלבים: *decreaseKey* שעולמו $O(log n)$ במקרה הגרוע, ו- *deleteMin*. בהנחה שהעירימה מוגדרת כל-עצלה (*lazyMelds=false*), גם *deleteMin* חסומה ב- $O(log^2 n)$ אם *lazyDecreaseKeys=false* אחרת חסומה ב $(n \cdot O(log n))$. בפרט, ב-W זמן הריצה הוא $(n \cdot O(n))$.
- meld** - הfonקציה ממזגת את העירימה הנוכחיית עם עירימה נוספת (*heap2*). תחילת הfonקציה מרשרת את רשימת השורשים של העירימה השנייה לקצה רשימת השורשים של העירימה הנוכחיית. לאחר מכן היא מעדכנת את השדות הסטטיסטיים (*size*, *numTrees*, *numMarkedNodes*) ואת המינימום הגלובלי. לבסוף, נבדק את הערך של *lazyMelds*: אם הוא *false*, מטבחת קירה ל- *successiveLinking*.

ניתוח סיבוכיות: מתחלך לפי ערכו של השדה *lazyMelds*:

 - אם *lazyMelds=true*: מטבח שרשור מצביעים בלבד ועדכו שdots, ולכן הזמן הוא $(1 \cdot O(n))$.
 - אם *lazyMelds=false*: מטבח קירה ל- *successiveLinking* הסורקת את כל השורשים. במקרה הגרוע, מספר השורשים פרופורציונלי ל-*n*. לכן, הזמן הריצה הוא $(n \cdot O(n))$.
- successiveLinking** - פונקציית עזר המבצעת את תהליך *Consolidation*. הfonקציה אחראית על ארגון מחדש של רשימת השורשים כך שלא ישארו שני עצים בעלי אותה דרגה. היא יוצרת מערך עזר בגודל לוגריתמי, מתחילה אותו ב-*null*, וקוראת לפונקציה *consolidate*. לאחר מכן היא מעדכנת את מצביעי ה-*head* וה-*last* בהתאם לרשימה החדשה שהתקבלה.

ניתוח סיבוכיות: הfonקציה תלואה ישירות בfonקציה *consolidate* הסורקת את כל שורשי העירימה. בפרט, במקרה הגרוע כמות השורשים לינארית במסות האיברים ולכן זמן הריצה הוא $(n \cdot O(n))$.
- link** - פונקציית עזר המחברת שני עצים בעלי דרגה זהה. הפעולה כוללת עדכון מצביעי הורם, ילדים ואחים, והגדלת הדרגה של השורש בעל המפתח הגדול.

ניתוח סיבוכיות: הfonקציה מבצעת מספר קבוע של שני מצביעים ופעולות אРИתמטיות. זמן הריצה הוא $(1 \cdot O(n))$.
- toBucket** - פונקציית עזר המעבירת את כל השורשים לtower מערך ה"סלים" (*Buckets*) ובמבצעת מיזוגים Tower CDI. הfonקציה עוברת על רשימת השורשים. עבור כל שורש, היא בודקת אם קיימں כבר עז באותו הדרגה ב"סל" המתאים. אם כן, היא מבצעת *link* ביניהם, מפנה את הסל הנוכחי, ומנסה להכניס את העז המאוחד לסל בדרגה הבאה (תהליך חוזר עד שנמצא סל ריק).

ניתוח סיבוכיות: הfonקציה עוברת על כל השורשים ברשימה המקורית. במקרה הגרוע כאמור עבור עירימות עצילות, יתכונו $(n \cdot O(n))$ שורשים ועל כן זמן הריצה יהיה $(n \cdot O(n))$, אחרית זמן הריצה יהיה $(log n)$.
- fromBucket** - פונקציית עזר הבונה מחדש את רשימת השורשים מתוך מערך הסלים. הfonקציה סורקת את מערך הסלים, וכל עז שנמצא בו מחובר לשרשראת חדשה של שורשים. במקביל, היא מעדכנת את מונה העצים (*numTrees*).

ניתוח סיבוכיות: הfonקציה עוברת על מערך הסלים שגודלו חסום לוגריתמית ומבצעת פעולות קבועות לכל תא. לכן, זמן הריצה הוא $(log n)$.
- consolidate** - פונקציה המאחדת את תהליך הסידור: קירה ל-*toBucket* לפיזור ומייזוג העצים, ולאחריה קירה ל-*fromBucket* לאיסוף העצים לרשימה חדשה.

ניתוח סיבוכיות: הפונקציה מבצעת את שני השלבים לעיל. השלב הדומיננטי הוא *toBucket* ועל כן בזמן הריצה הכלול בWC יהיה $O(n)$ עבור עריםות עצלות, אחרת ($O(logn)$).

- **getters** - הפונקציות הבאות מחזירות ערכים של שדות המחלקה. כל הערכים מתוחזקים באופן שוטף במהלך פעולות העירמה, ולכן השליפה היא מיידית בזמן ריצה של $O(1)$.
 - *size* –מחזירה את מספר האיברים בעירמה.
 - *numTrees* –מחזירה את מספר העצים (השורשים) בעירמה.
 - *numMarkedNodes* –מחזירה את מספר הצמתים מסומנים.
 - *totalLinks* –מחזירה את סך פעולות הקישור שבוצעו.
 - *totalCuts* –מחזירה את סך פעולות החיתוך שבוצעו.
 - *totalHeapifyCosts* –מחזירה את כמות הפעולות שבוצע תחילה ה-*heapify*.

חלק ניסויי/תיאורטי

1. נגדיר פונקציית פוטנציאלי. בוגוד לרימות פיבונאצ'י רגילה, שבה פונקציית הפוטנציאלי היא $m + T$, כאן מספר העצים תמיד חסום ע"י ($O(logn)$) בגלל של פעלות שגוראת ל*meld* מבצעת *successive linking*. כמו כן, מואתיה סיבה כל ניתוק עולה ($O(logn)$). לכן נגדיר את פונקציית הפוטנציאלי להיות $m * logn + T = O(logn + m * logn)$.

אחריה *actual cost* – *insert* ($O(logn)$) ו- *successive linking* ($O(1)$) ו- *actual cost* – *findMin* ($O(logn)$). כמו כן, $(T_0 + m * logn) - (T_1 + m * logn) = O(1)$. לכן סה"כ עלות בנייה לשיעורין היא $O(logn)$.

- **deleteMin** – נסיק כי שארינו בהרצאה, *actual cost* – *deleteMin* ($O(logn)$) נקבע כי מתקיים במקרה זה $T_0 + logn \leq T_1 + m_1 * logn$. לכן $\Delta\Phi = T_1 + m_1 * logn - (T_0 + m_0 * logn) = m_1 - m_0$. מכיוון ש-*amortize* ($O(logn)$) מוגדר כ-*actual cost* – *insert* ($O(logn)$) ועוד ש-*logn* ($O(logn)$) מוגדר כ-*actual cost* – *successive linking* ($O(1)$), אז $\Delta\Phi = O(logn)$.

- **decreaseKey** – במקרה זה, *actual cost* – *decreaseKey* ($O(clogn)$) כאשר c הוא כמות החיתוקים שמבצעים בפעולה, זאת מאחר וכל חיתוך גורר הוספה לרשימה וביצוע *meld* נפרד (אומרו, *meld* ($O(1)$) מבנית ניתוח השינוי בפוטנציאלי, ראשית נשים לב כי שארינו בהצראה כי השינוי בכמות c ($O(logn)$) מוגדר כ-*actual cost* – *decreaseKey* ($O(clogn)$)). לכן קיבלנו כי השינוי בפוטנציאלי הוא $c - 1 = 2 - c$. מכיוון ש-*logn* ($O(logn)$) מוגדר כ-*actual cost* – *successive linking* ($O(1)$), אז $\Delta\Phi = \Delta T + \Delta m * logn = c + (2 - c)logn = 2logn$. מכיוון נקבע כי סה"כ ניתוח של זמן הריצה לשיעורין הוא $O(logn)$, *amortize* ($O(logn)$) מוגדר כ-*actual cost* – *decrease key* ($O(clogn + c + 4logn - clogn) = 4logn + c = O(logn)$).

- **delete** – קוראת ל- *delete min* ו-*decrease key* ($O(logn)$). הראמנו כי ניתוח לשיעורין של שתי הפעולות הוא *amortize* ($O(logn)$). לכן סך הכל $O(logn)$.

.2

פעולה	ערימה ביןומית ערך	ערימה פיבונאצ'י ערך	ערימת פיבונאצ'י ערך	ערימה ביןומית ערך	ערימה ביןומית ערך
<i>Insert</i>	$O(1)$	$O(1)$	$O(1)$	$O(1)$	$O(1)$
<i>findMin</i>	$O(1)$	$O(1)$	$O(1)$	$O(1)$	$O(1)$
<i>deleteMin</i>	$O(logn)$	$O(logn)$	$O(logn)$	$O(logn)$	$O(logn)$

$O(logn)$	$O(1)$	$O(logn)$	$O(logn)$	decreaseKey
$O(logn)$	$O(logn)$	$O(logn)$	$O(logn)$	delete

.3

ניסוי 1	ערימה בינומית עם ניתוקים	ערימת פיבונacci	ערימה בינומית עצלה	ערימה בינומית
$O(nlogn)$		$O(n)$	$O(n)$	$O(n)$
$O(nlogn)$		$O(nlogn)$	$O(nlogn)$	$O(nlogn)$
$O(nlogn)$		$O(n)$	$O(nlogn)$	$O(nlogn)$

.4

ניסוי 1:

זמן ריצה (מילישניות)	ערימה בינומית עם ניתוקים	ערימת פיבונacci'	ערימה בינומית עצלה	ערימה בינומית
116	18		19	123
464645	464645		464645	464645
9	9		9	9
464653	464653		464653	464653
0	0		0	0
0	0		0	0
18	464636		464636	18
עלות מקסימלית לפעולה				

זמן ריצה (מילישניות)	ערימה בינומית עם ניתוקים	ערימת פיבונacci'	ערימה בינומית עצלה	ערימה בינומית
505	209		888	1205
46	46		46	46
2	2		4	4
1405996	748761		7552691	7552819
748756	748717		0	0
מספר חיטוכים				

0	0	4148715	4149301	סה"כ עליות של <i>heapify up</i>
39	464636	464636	35	עלות מקסימלית לפעולה

ניסוי 2:

ניסוי 3:

ערימה בינומית עם ניתוקים	ערימת פיבונacci	ערימה בינומית עצלה	ערימה בינומית	סה"כ עליות של <i>heapify up</i>
121	24	23	115	זמן ריצה (מילישניות)
464644	464644	464644	464644	גודל הערימה בסיום
11	11	8	8	מספר העצים בסיום
509262	509247	464653	464670	מספר חיבורים
44610	44614	0	0	מספר חיתוכים
0	0	114600	114559	סה"כ עליות של <i>heapify up</i>
18	464636	464636	18	עלות מקסימלית לפעולה

.5

א. אכן ניתן לראות כי קיימים הבדלים ביצועים של הערים בניסויים השונים.

ניסוי הראשון, ניתן לראות כי הערים אשר $\text{meld} = \text{True}$, (*פיבונacci* ועצלה) הפגינו ביצועים טובים יותר. זאת מכיוון ש *meld lazy* חוסכת זמן על ידי דחית איחוד העצים לפעולות החום *min delete*. לעומת זאת, הערים האחרות מבצעות את האיחודים בכל פעולה *Insert* ומכאן נובע הפער בזמן הריצה בין שני סוגי הערים.

ניסוי השני, נשים לב כי הערים אשר $\text{decrease keys} = \text{True}$, (*פיבונacci* ובינומית עם ניתוקים) הציגו ביצועים טובים יותר. זאת מכיוון שפעולות *key decrease* מתחכמת ב(1) לשינויו לעוממת הערימה הבינומית והערימה העצלה שבן מבצעים את פעולה *heapify up* שלוקחת (*logon*) לשינויו. ניתן לראות בטבלה כי מספר החיתוכים שביצעו בערים (בערך 750,000) קטן בהרבה ממספר פעולות *heapify up* שביצעו (בערך 4,150,000) ובנוסף לכך שפעולה זאת יקרה בהרבה גורם להבדל הגדול ביצועים.

בנוסף גם כאן ניתן לראות הבדלים בין הערים אשר ביצעו *meld lazy* הנובעים מהחלק הראשון של הניסוי זהה לניסוי 1.

ניסוי השלישי, באופן דומה לניסוי 1, הערים עם הביצועים הטובים יותר הם *פיבונacci* ועצלה. זאת מכיוון שמספר החיתוכים, *up heapify* זניח ביחס לכמות ה *Linking* כפי שניתן לראות בטבלה, لكن הערים עם *lazy meld* הציגו ביצועים טובים יותר.

ב. תוצאות הניסוי בסעיף 4 מאשרות את התוצאות בסעיף 3.

ראשית, נשים לב שלפי ההוראות, מימוש הקוד לערימה בינהית מתנהג כמו עירימה בינהית אבל הביצועים שלו אינם זהים לערימה בינהית. لكن בניסויים שביצענו זמני הריצה של עירימה בינהית גבוהים מהצפוי. בניסוי הראשון אכן רأינו כי עירימות פיבונאצ'י ועכלה היו מהירות יותר כפי שציינו בסעיף 3. כמו כן רأינו כי עירימה בינהית עם ניטוקים הגיעו לביצועים הנמוכים ביותר. לעומת זאת זמני הריצה הנמוכים של עירימה בינהית בניסוי נובעים מהמיימוש בקוד וכן אין תואמים את סעיף 3.

בניסוי השני אכן הגיעו לסדרי גודל דומים יותר בין העירימות ובניסוי השלישי כפי שצפינו עירימת פיבונאצ'י הגיעה לביצועים הטוביים ביותר (עם הסתיגות על בינהית עצלה).

ג. ניתן לראות כי בכל הניסויים הוללות המקסימלית לפעולה בפיבונאצ'י ועכלה היא (ח) 0 לעומת בינהית ובינהית עם ניטוקים שם הוללות המקסימלית לפעולה קטנה משמעותית. זאת מכיוון שבירימות עם lazy True = meld פעולה המכילה הראשונה הינה הפעולה הכבודה ביותר והיא משלמת על כל הפעולות הזרלות שביצענו עד כה. לעומת זאת, זמני הריצה שלhn קצרים יותר משל שתי האחרות, שכן הוללות לשיעורן שלhn נמוכה יותר כפי שנלמד בהרצאה.