מבני נתונים
תרגיל 5 - פתרונות

גלעד אשרוב

3 ביולי 2014

שאלה 1. נניח שלמונה הבינארי שהוגדר בתרגול הוספנו את פעולת ה$decrement$ (להחסיר ב-1). הוכיחו או הפריכו את הטענה הבאה: העלות לשיעורין של פעולות כלשהו על המונה הבינארי הוא $O(1)$.

פתרון. הטענה אינה נכונה. מספיק להראות קיום של סידרה שעבורה מקבלים זמן ריצה גבוה מ- $O(1)$ בממוצע לכל פעולה. הסידרה תהיה מורכבת מהמון פעולות $increment$ עד שנגיע למצב שבו הביט הראשון במונה דלוק, וכל שאר הביטים מכובים (כלומר, למצב $1000\ldots$). בשלב זה, נתחיל לבצע רצפים של זוגות $decrement, increment$. ברור שכל זוג כזה מחליף את כל הביטים, ולכן כל פעולה עולה k. ישנם 2^k פעולות עד שמגיעים למצב $100\ldots$, שעלותן היא קטנה מ - $2 \cdot 2^k = 2^{k+1}$ (כפי שראינו בכיתה), אבל היא ודאי גדולה מ: 2^k (שכן עלות כל פעולה היא לפחות 1). אם כן, עבור סידרה של נניח, 2^{k+1} פעולות, נקבל כי:

$$T(2^{k+1}) \geq 2^k \cdot 1 + 2^k \cdot k \geq 2^k \cdot k$$

ולכן, עלות הממוצעת לכל פעולה היא:

$$\frac{T(2^{k+1})}{2^{k+1}} \geq \frac{2^k \cdot k}{2^{k+1}} \geq \frac{k}{2}$$

.

שאלה 2. בתרגול למדנו מערך דינאמי הגדל פי שניים בכל פעם, והתייחסנו לפעולת הכנסה בלבד. כעת, נתבונן במקרה בו ייתכנו פעולות הכנסה והוצאה (כאשר ההכנסה / הוצאה מתבצעת על האיבר האחרון במערך). נניח שכעת אנו במצב של מערך בגודל n. כאשר נגיע לתפוסה מלאה במערך (מספר האיברים במערך הוא n) - נקצה מערך בגודל $2n$ ונעתיק אליו את n האיברים. מן הצד השני, כאשר במערך רק $n/4$ איברים, נקצה מערך חדש בגודל $n/2$, ונעתיק את הערכים למערך החדש (עלות - $n/4$).

מצא חסם לעלות של רצף של m פעולות, לכל $m \geq 0$. (מומלץ לבצע ניתוח לפי שיטת "הבנק", אך אין הגבלה על השיטה).

פתרון: ננתח לפי שיטת הבנק. עם כל הכנסה - נשלם 3 יחידות כפי שעשינו במקרה הרגיל. עם כל הוצאה, נוציא שתי יחידות מהכיס - אחת עבור העלות האמיתית, ועוד יחידה נפקיד לבנק. נקבל אם כן (נניח כרגע מערך מגודל n ואנחנו מיד לאחר הרחבה, כלומר, $n/2$ איברים מלאים):

- אם אנחנו מכניסים עוד $n/2$ אלמנטים, נקבל כי בבנק יש $n/2 \cdot 2 = n$ יחידות, וזוהי בדיוק העלות להקצאת מערך חדש.
- אם אנחנו מורידים $n/4$ אלמנטים, נקבל כי בבנק יש בדיוק $n/4$ יחידות, וזוהי בדיוק העלות להקטנת המערך.

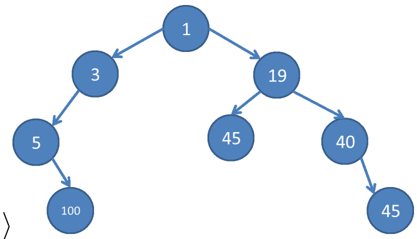
כל אפשרות אחרת היא בעצם הכנסת איבר והוצאתו או הוצאת איבר והכנסתו. בשני המקרים - נישאר עם תוספת חיובית בבנק מעבר למה שאנחנו באמת צריכים. לכן, נסכם כי לעולם לא ניכנס למינוס בבנק. נקבל אם כן:

$$\sum_{i=1}^{m} c_i \leq \sum_{i=1}^{m} \hat{c}_i \leq 3m$$

שאלה 3. בהינתן סידרה של n מספרים $A = (a_1, \cdots, a_n)$, העץ הקרטזי של A מוגדר בצורה רקורסיבי בצורה הבאה. אם $n = 0$, אזי העץ הוא עץ ריק. אחרת, יהי i המיקום של האיבר המינימלי ב - A. כלומר, לכל $1 \leq j \leq n$, $j \neq i$, $a_j \geq a_i$. השורש של העץ הקרטזי של A הוא צומת r עם הערך a_i, הבן השמאלי של r הוא העץ הקרטזי עבור הסידרה $A_{\text{left}} = (a_1, \ldots, a_{i-1})$, והבן הימני של r הוא העץ הקרטזי $A_{\text{right}} = (a_{i+1}, \cdots, a_n)$.

1. צייר עץ קרטזי עבור הסידרה $A = (5, 100, 3, 1, 45, 19, 40, 45)$.

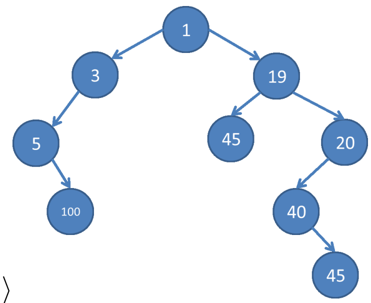
פתרון:



$\langle\rangle$

2. נניח ואנו רוצים להוסיף את האיבר 20 לסוף הסידרה הנ"ל. כיצד נראה העץ הקרטזי החדש?

פתרון:



$\langle\rangle$

3. מהו גובהו של העץ הקרטזי במקרה הגרוע ביותר?

פתרון: במקרה הגרוע ביותר, ייתכן ונקבל עץ שרוך ולכן הגובה הוא $O(n)$. מקרה זה מתקבל, לדוגמא, כאשר הקלט הוא סידרה ממויינת בסדר עולה.

4. הצג אלגוריתם המקבל כקלט עץ קרטזי עבור הסידרה $A = (a_1, \ldots, a_n)$ ומספר k, ומחזיר עץ קרטזי עבור הסידרה $A' = (a_1, \ldots, a_n, k)$. מהי סיבוכיות זמן הריצה של האלגוריתם שהצגת?

הנח: בייצוג העץ הקרטזי - לכל צומת קיים מצביע לבן הימני, לבן השמאלי ולאבא. בנוסף, אנחנו מחזיקים מלבד מצביע לשורש גם מצביע לאיבר האחרון a_n.

פתרון: המיקום של האיבר החדש יהיה בשדרה הימנית של העץ. בעזרת המצביע לאיבר a_n, נעלה בשדרה הימנית (בעזרת המצביע ל"אבא") עד לאיבר הראשון שקטן ממש מהאיבר k, נניח איזשהו a_i. נציב את k כבנו הימני של a_i, ונציב כבנו השמאלי של k את בנו הימני המקורי של a_i. נשים לב שלמעשה k הוא האיבר הכי ימני בעץ, ונמצא בקצה השדרה הימנית. בדקו שהינכם מבינים מדוע האלגוריתם עונה על הדרישה.

עלות האלגיריתם היא כמות האיברים הגדולים מ־k בקצה השדרה הימנית של העץ.

5. נניח את האלגוריתם הבא לבניית עץ קרטזי עבור הסידרה $A = (a_1, \ldots, a_n)$: מתחילים מעץ ריק המתאים לסידרה $A_0 = ()$. בכל שלב, אנחנו מפעילים את אלגוריתם ההכנסה ("השרשור") מהסעיף הקודם, ומוסיפים את האיבר הבא בסידרה. כלומר, בסיום השלב ה - i בנינו עץ קרטזי עבור הסידרה $A_i = (a_1, \ldots, a_i)$.
בשלב ה־$i+1$ נכניס את האיבר ה - a_{i+1} לעץ, ונקבל עץ קרטזי שמתאים לסידרה $A_{i+1} = (a_1, \ldots, a_{i+1})$.
בסיום השלב ה-n, אנחנו מסיימים עם עץ קרטזי עבור $A_n = A$.

הראה שהעלות לשיעורין לכל הכנסה הוא $O(1)$. במילים אחרות, הראה שבניית עץ קרטזי עבור הקבוצה A הוא במקרה הכי גרוע $O(n)$.

רמז: פתור בעזרת שיטת הפוטנציאל. הגדר את הפונקציה להיות אורך המסלול מהאיבר שהוכנס אחרון לשורש העץ.

פתרון: נשים לב שעלות כל איטרציה ה־i היא כמות האיברים בשדרה הימנית התחתונה עד לאיבר שקטן מהאיבר a_i.

נשים לב לעובדה הבאה: אם בפעולה הנוכחית אנחנו צריכים לשלם "הרבה" (כלומר, לעלות הרבה רמות), פוטנציאל עלות הפעולה הבאה **קטן**, שכן ישנם פחות איברים בשדרה הימנית. לעומת זאת, אם לא עלינו כלל - כמות האיברים בשדרה הימנית גדל, ולכן פוטנציאל עלות ההכנסה הבאה - גדל אף הוא.

למעשה, כאשר נגדיר פונקציית פוטנציאל לפי הרמז (נשים לב שבמקרה ההתחלתי הפוטנציאל הוא 0 ולכל מצב אחר הפוטנציאל אי־שלילי, ולכן לכל i מתקיים למעשה $\phi(D_i) \geq \phi(D_0) = 0$ והפונקציה "חוקית"), ערך הפונקציה בכל שלב יהיה בדיוק אורך השדרה הימנית של העץ. עם כל הכנסה שבה האיבר נכנס רק לסוף השדרה הימנית ־ הפוטנציאל גדל באחד, ובכל הכנסה שבה האיבר עולה מספר רמות ־ הפוטנציאל קטן כמספר הרמות שעלה, וכמו כן אורך השדרה הימנית. התשלום במקרה הזה יבוצע על ידי פירוק הפוטנציאל שנצבר בפעמים הקודמות.

נקבל אם כן שבכל הכנסה שבה לא עולים רמה ־ נשלם שניים (אחד עבור ההכנסה + אחד עבור גידול הפוטנציאל), ולכל הכנסה שבה עולים רמה ־ נשלם 0 (נשלם ע"י פירוק בפוטנציאל מהשלבים הקודמים). לפיכך, עלות ההכנסה היא $O(1)$ לשיעורין, ולכן עלות בניית העץ כולו היא $O(n)$ במקרה הגרוע.

שאלה 4. עליכם לבנות מבנה נתונים התומך בפעולות הבאות:

- הכנסה (תעשה תמיד לאינדקס הראשון שפנוי במבנה).
- גישה לאיבר באינדקס-i (עלות נדרשת - $O(1)$).
- אין צורך לתמוך בהוצאה.

עליכם לבנות מבנה נתונים התומך בפעולות הנ"ל. המטרה ־ להביא למינימום את עלות ההכנסה, ולהביא למינימום את כמות הזיכרון שלא מנוצל.

1. הראו מבנה נתונים התומך בנ"ל כך שעבור כמות זיכרון לא מנוצל - k, עלות ההכנסה היא $O(n/k)$ לשיעורין.

פתרון: פשוט מערך דינאמי שבו מגדילים את המערך ב־k כל פעם ולא פי שניים כפי שראינו בכיתה. אכן, כאשר $k = 1$ נקבל שעלות הכנסה היא $O(n)$ (צריך להעתיק את כל המערך מחדש בכל הכנסה), וכאשר

$k = n$ נשלם $O(1)$ להכנסה (כפי שראינו במערך דינאמי ללא ההגבלה שבשאלה). באופן כללי, נצטרך לבצע ההעתקה $O(n)$ כל k הכנסות, ולכן בממוצע עלות הכנסה היא $O(n/k)$.

2. הראו מבנה נתונים כך שעבור כמות זיכרון לא מנוצל $c \cdot k$ (עבור c קבוע כלשהו), עלות ההכנסה היא $O(n/k^c)$ לשיעורין.

פתרון: נפתור תחילה עבור $c = 2$. במקרה זה יהיו לנו שתי רמות. ברמה הראשונה נשמור מערך של מצביעים למערכים בגודל k כל אחד. בהתחלה, יהיה לנו רק מערך של מצביעים (באורך k) המצביע למערך יחיד בגודל k. לאחר k הכנסות, נקצה מערך חדש בגודל k והאיבר השני במערך המצביעים יצביע לאותו המערך. לאחר k^2 הכנסות, מערך המצביעים יהיה מלא, ונצטרך להעתיק אותו למערך חדש. בכדי לגשת לאיבר ה - i, נחשב i/k ונקבל באיזה מערך הוא נמצא. ניגש למערך המצביעים באינדקס הנ"ל ונקבל מצביע למערך בגודל k שבו האיבר נמצא. האיבר יימצא שם באינדקס $i \bmod k$.

למעשה, בצורה כזו, אנו נדרשים לבצע ההעתקה כל k^2 הכנסות, ולפיך נשלם $O(n)$ כל k^2 הכנסות, מה שאומר שהעלות הממוצעת לפעולה היא $O(n/k^2)$, כאשר בכל שלב במהלך הריצה אין לנו יותר מ־$2k$ מקום לא מנוצל.

עבור $c = 3$, יהיה לנו ברמה ראשונה מערך של מצביעים לאיברים ברמה השנייה - שיהיו גם הם מצביעים, לרמה שלישית - שבה באמת ישבו האיברים. ההעתקה תתבצע רק עבור הרמה הראשונה, ונצטרך לבצע ההעתקה רק לאחר k^3 הכנסות. לפיכך, העלות לשיעורין של כל פעולה היא $O(n/k^3)$. המעבר ל - c כללי (קבוע) הוא ברור.

שאלה 5.

1. נתונים שני עצי $B\text{–}tree$, A ו־B מסדר m. העץ A מכיל a איברים והעץ B מכיל b איברים. נניח $a > b$, ונניח שכל האיברים ב־A גדולים מכל האיברים ב־B. נסמן ב־h את גובהו של העץ A. תארו אלגוריתם הבונה $B\text{–}tree$ מאוחד, גם הוא מסדר m בזמן $O(h)$.

2. נתון עץ $B\text{–}tree$ מסדר m, ונתון איבר x. נסמן ב-h את גובהו של העץ. תארו אלגוריתם המפצל את העץ לשני עצי $B\text{–}tree$, A ו־B מסדר m כך שכל האיברים בעץ A קטנים שווים ל־x, וכל האיברים בעץ B גדולים מ־x. האלגוריתם צריך לעבוד בזמן $O(h)$.

פתרון: לא יפורסם פתרון לתרגיל זה.