תרגול 8 - עצי AVL

28.4.14

עצי AVL

עצי חיפוש בינארי שמקיימים את התנאי שעבור כל קודקוד A גובה תת העץ הימני שונה בלכל היותר 1 מגובה תת העץ השמאלי.

$$|h(A_L) - h(A_R)| \leq 1$$

אנחנו נניח בה"כ שאין ערך שמופיע פעמיים.
נוכיח שגובה עץ AVL הוא לוגריתמי.
נסמן ב n_k את מספר הקודקודים המינימלי בעץ AVL מגובה k.

טענה 0.1 $n_k = n_{k-1} + n_{k-2} + 1$ ובפרט n_k מונוטוני עולה ממש ב k, כלומר $n_k < n_{k+1}$.

הוכחה: לשורש עץ מגובה k יש בן שהוא שורש עץ מגובה $k-1$. בתת עץ זה יש ממינימליות n_{k-1} קודקודים. ביחד עם השורש נקבל ש $n_{k-1} < n_k$.
כעת מהמונוטוניות נקבל שהבן השני חייב להיות n_{k-2}. ■

$(n_k + 1) = (n_{k-1} + 1) + (n_{k-2} + 1)$. כמובן ש $n_0 = 1\,,\ n_1 = 2$ ואכן $n_2 = 4$.
נגדיר $m_k = n_k + 1$ אזי $m_k = m_{k-1} + m_{k-2}$, אזי $m_k = m_{k-1} + m_{k-2}$, $m_0 = 2,\ m_1 = 3$.

פיבונאצ'י

הגדרה 0.2 נגדיר באינדוקציה סדרת מספרים:

- $F_0 = 0\,,\ F_1 = 1$
- $\forall n > 1\ F_n = F_{n-1} + F_{n-2}$

תחילת הסדרה נראית כך: $0, 1, 1, 2, 3, 5, 8, 13, 21, 34$.

למה 0.3 $2^{\left[\frac{n}{2}\right]} \leq F_n \leq 2^{n-1}$

הוכחה: $2F_{n-2} \leq F_n = F_{n-1} + F_{n-2} \leq 2F_{n-1}$ ■

מסקנה 0.4 $n_k = \Omega\left(\sqrt{2}^k\right)$, $n_k = O\left(2^k\right)$ ולכן $k = \Theta(\log n_k)$ ו $k = O(\log n)$.

אפשר למעשה לקבל חסם הדוק לגובה העץ:

$$k \leq 1.44 \log n_k + O(1)$$

בעזרת הטענה הבאה: (לא נעביר בתרגול)

טענה 0.5 $\forall n \geq 0\ F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right]$

הוכחה: מהלמה ל F_n יש קצב גידול מעריכי, נציב $F_n = x^n$ אזי $F_n = x^{n-1} + x^{n-2}$ ומכאן $x^2 = x + 1$ ולכן $x_{1,2} = \frac{1 \pm \sqrt{5}}{2}$.
אכן $g_n = x_1^n$ מקיים $g_n = g_{n-1} + g_{n-2}$ וכנ"ל ל $h_n = x_2^n$, אך לא את תנאי ההתחלה.
נגדיר $k_n = \alpha g_n + \beta h_n$ אזי $k_n = k_{n-1} + k_{n-2}$.
נפתור $k_0 = 0$ ו $k_1 = 1$ ונמצא ש $\alpha = \frac{1}{\sqrt{5}}, \ \beta = -\frac{1}{\sqrt{5}}$ ומכאן הטענה. ■

נחשב קירוב:
כיוון ש $\frac{1}{2} > \frac{1}{\sqrt{5}} \cdot \left|\frac{1-\sqrt{5}}{2}\right|^n$ $\forall n$ נקבל שפשוט

$$F_n = \left[\left(\frac{1+\sqrt{5}}{2}\right)^n \cdot \frac{1}{\sqrt{5}}\right] + O(1) = \Theta(\varphi^n)$$

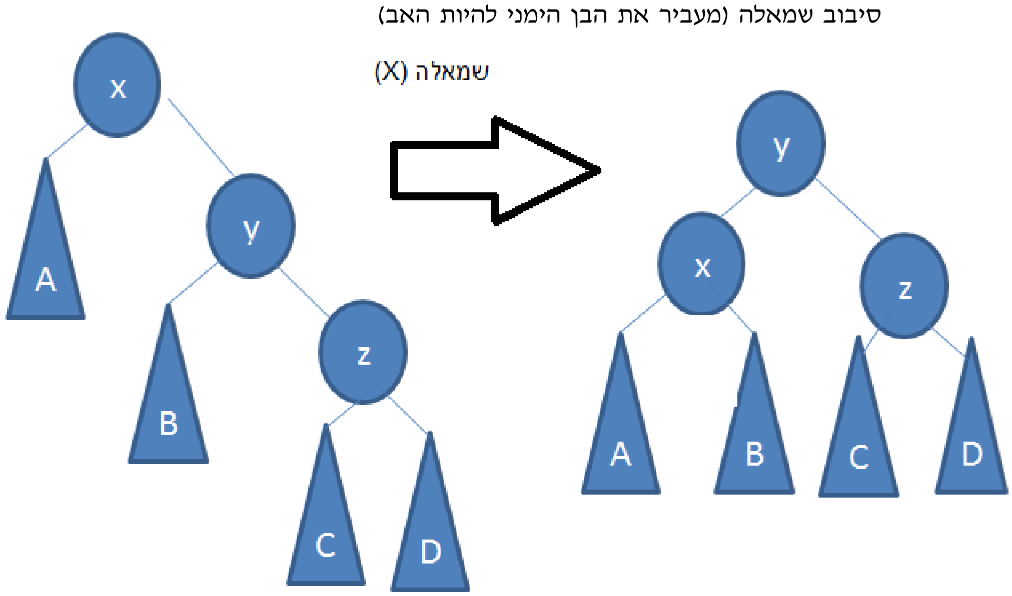
עבור $\varphi = \frac{1+\sqrt{5}}{2} \sim 1.618$
כעת $m_k = F_{k+3}$ ולכן $\left[\left(\frac{1+\sqrt{5}}{2}\right)^{k+3} \cdot \frac{1}{\sqrt{5}}\right] \geq n_k = F_{k+3} - 1$ כלומר $\log n_K \geq$ $(k+3)\log\varphi + O(1)$ ומכאן

מסקנה 0.6 $k \leq \frac{\log n_k}{\log \varphi} + O(1) \sim 1.44 \log n_k + O(1)$.

הכנסת איבר

הגדרה 0.7 נגדיר הפרת LL כמקרה שבו יש הפרה בקודקוד A והמסלול באורך המירבי מ A לעלה עובר שמאלה ושמאלה מ A.

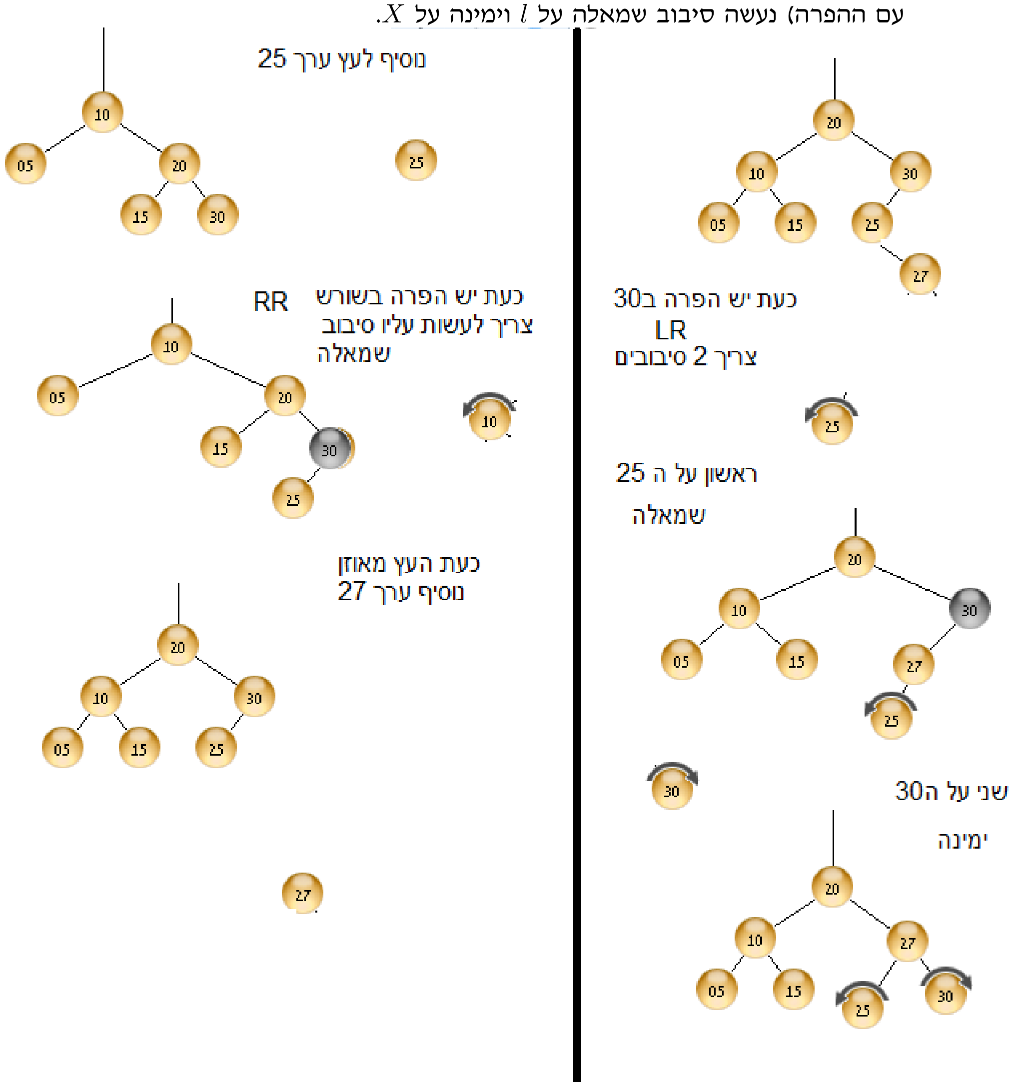
נכניס איבר באופן רגיל ונבדוק הפרה של האיזון.
יש 4 מקרים. מטפלים בהם באמצעות סיבובים:
סיבוב שמאלה (מעביר את הבן הימני להיות האב)



ימינה זה סימטרי.

RR: אם האיבר שהוכנס נמצא בתת העץ הימני של הבן הימני של X (הקודקוד עם ההפרה) נעשה סיבוב שמאלה X.

LR: אם האיבר שהוכנס נמצא בתת העץ הימני של הבן השמאלי l של X (הקודקוד עם ההפרה) נעשה סיבוב שמאלה על l וימינה על X.



מחיקת איבר

ראשית נמצא את האיבר x אותו רוצים למחוק.

- אם ל x יש שני בנים:
  כאמור מחיקה מתבצעת ע"י מציאת העוקב y, החלפת ערך x ב y, ומחיקת y, שלו יש בן אחד לכל היותר - ומקרה זה מטופל.

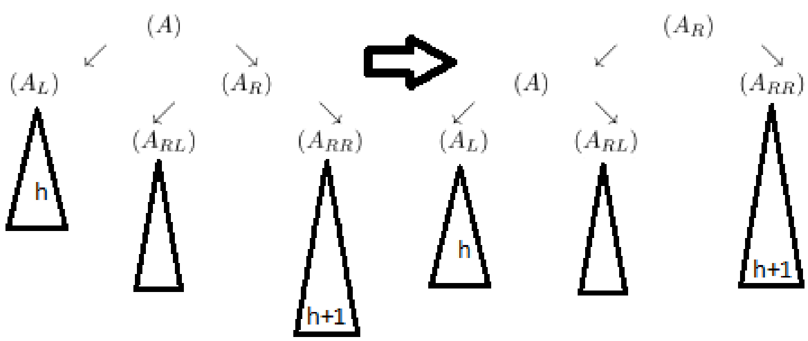
- אחרת ־ נבצע את המחיקה כמו ב BST ונעדכן את הגבהים ־ לכיוון השורש.
נשים לב שלא יכולה להיות הפרה בתתי עצים של x, אלא רק במסלול של x לשורש.
נתחיל לעלות למעלה מהקודקוד לשורש, עד אשר נגיע בפעם הראשונה להפרה.
אותו טיפול כמו בהכנסה יפתור גם מחיקה.
כעת צריך לחפש מיהו העלה העמוק ביותר מנק' זו (בניגוד להכנסה). אנחנו לא יודעים מיהו העלה, אלא רק באיזה תת עץ הוא ביחס להפרה (תת העץ שלא בו בוצעה המחיקה). ואז נדע איזה סוג הפרה יש לנו. אותו טיפול בהפרות כמו בהכנסה יסדר את העץ.

טענה 0.8 אם נוצרה הפרה RR בקודקוד A בעקבות מחיקה, אזי סיבוב שמאלה על A יתקן את ההפרה בתת העץ של A.

הוכחה: אם ההפרה היא RR אזי בהכרח מחקנו איבר שנמצא ב A_L.

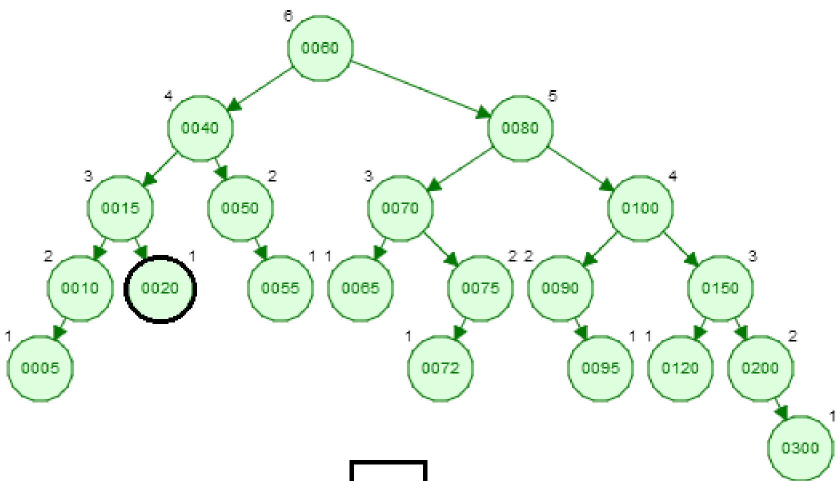
- אם (בשלילה) האיבר שמחקנו היה ב A_{RL}, (ולא הייתה הפרה ב A_R), אז נסמן ב h את גובה A_{RR}, אז לפני המחיקה $h\left(A_{RL}\right) = h+1$ ואחריו זה היה h, ו $h\left(A_L\right) = h+2$, ולכן המקרה הוא לא RR.

- אם האיבר שמחקנו היה ב A_L, אזי נסמן ב h את $h\left(A_L\right)$ אחרי המחיקה, אזי לפני המחיקה $h\left(A_L\right) = h+1$ ו $h\left(A_R\right) = h+2$.
כיוון שאנחנו ב RR, אזי $h\left(A_{RR}\right) = h+1$, ולכן $h \leq h\left(A_{RL}\right) \leq h+1$.
לאחר סיבוב שמאלה נקבל ש $h\left(A_{RR}\right), h\left(A_{RL}\right), h\left(A_L\right)$ לא השתנה כתוצאה מהסיבוב.
$h+1 \leq h\left(A\right) \leq h+2$ ולכן אין הפרה בו או בהורה שלו כעת A_R.

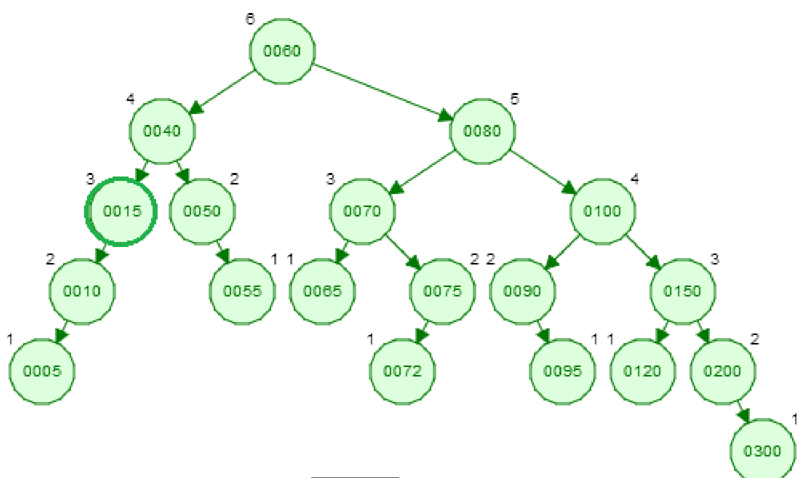


■

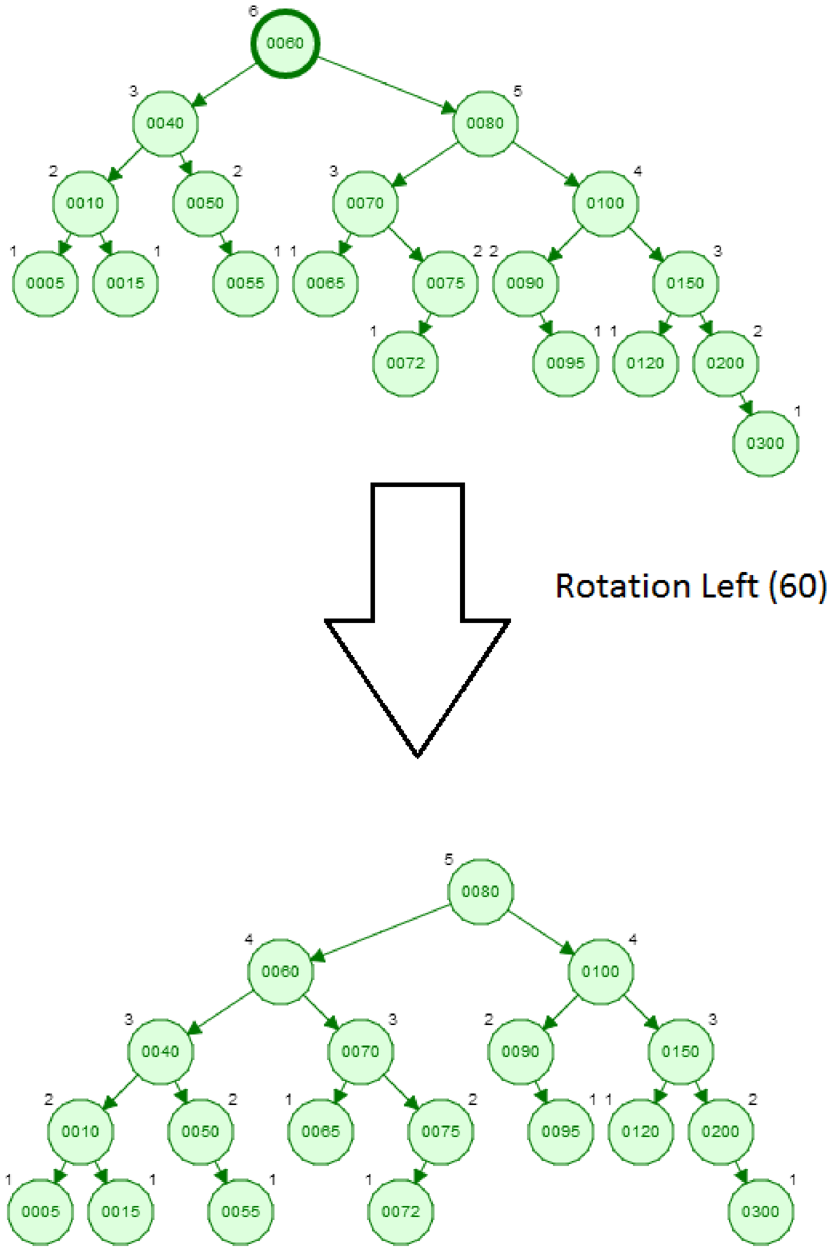
נשים לב שאמנם תיקנו את ההפרה בתת העץ של A (שהוא עכשיו תת העץ של A_R) אבל ייתכן עדיין שבהורה של תת העץ הנ"ל יש הפרה ־ ולכן צריך להמשיך לבדוק ־ עד שמגיעים לקודקוד שהגובה שלו לא השתנה בעקבות המחיקה (או עד השורש).



Delete 20



Rotation Right (15)


0060
0040
0080
0010
0050
0070
0100
0005
0015
0055
0065
0075
0090
0150
0072
0095
0120
0200
0300
Rotation Left (60)
0080
0060
0100
0040
0070
0090
0150
0010
0050
0065
0075
0095
0120
0200
0005
0015
0055
0072
0300

יצירת עצי AVL

נניח שיש לנו מערך ממוין באורך n ואנחנו רוצים לבנות ממנו עץ AVL.
אפשר באופן טריוויאלי לעבור איבר איבר ולהוסיף אותו לעץ ולדאוג שהוא יהיה AVL לאחר כל הוספה. זה יקח $\Theta(n\log n)$. למה? כי אחרי שנכניס מחצית האיברים - גובה העץ יהיה לוגריתמי (תכונת AVL) וכל הוספה של איבר מאותו רגע תעלה זמן $O(\log n)$ (תרגיל).
נרצה לבנות את העץ בזמן לינארי. הרעיון פשוט - נשים בשורש את האיבר האמצעי (חציון) ונקרא ברקורסיה לבן השמאלי עם חצי המערך הקטן ולימני עם החצי הגדול:

$\boldsymbol{BUILD\ AVL(A[1\dots n], start, end)}$
$N \leftarrow start - end + 1$
$if\ N < 0\ then\ return\ NIL$
$r \leftarrow start + \left\lfloor \frac{n-1}{2} \right\rfloor$
$tree \leftarrow new\ Node(A[r])$
$if\ N > 1$
  $left(A[r]) \leftarrow BUILD\ AVL(A, start, r-1)$
  $right(A[r]) \leftarrow BUILD\ AVL(A, r+1, end)$
$return\ tree$

זה לינארי - כי עוברים על כל קודקוד פעם אחת.
למה זה עובד?
העץ שמתקבל הוא BST - ברור - כי לכל קודקוד, הערכים של תת העץ השמאלי כולם קטנים ממנו, וכל ערכי תת העץ הימני גדולים ממנו.
העץ הוא AVL: נוכיח באינדוקציה. נשים לב שגובה העץ המתקבל לא תלוי בערכים במערך - אלא רק בגודלו והוא מונוטוני.
בסיס - עבור $n=1$ - ברור. נניח לכל $k<n$ ונוכיח ל n.
לשורש יש שני בנים, ומהנחת האינדוקציה כל אחד מהם הוא בעצמו AVL.
אם n א"ז אז גודל תת העץ השמאלי הוא $\frac{n-1}{2}$ וגם הימני. לכן זהו AVL.
אם n זוגי אז גודל תת העץ השמאלי הוא $\frac{n}{2}-1$ ואילו הימני הוא $\frac{n}{2}$. נסמן ב H את גובה העץ, H_r, H_l את גובה תת העצים. אזי $H_r \geq H_l$ ולכן $H = H_r + 1$. נשאר להראות ש $H_r \leq H_l + 1$.
מהנחת האינדוקציה עץ עם $n-2$ קודקודים הוא AVL. לו יש שני בנים עם תתי עצים בגדלים $\frac{n}{2}-1$ ו $\frac{n}{2}-2$ ולכן $H_{\frac{n}{2}-1} \leq H_{\frac{n}{2}-2}+1$ וגם $H_r \leq H_{n-1} = H_{\frac{n}{2}-1}+1$.