Lecture 4 of Geiger & Itai's slide brochure
www.cs.technion.ac.il/~dang/courseDS

עצי AVL

עצים מאוזנים

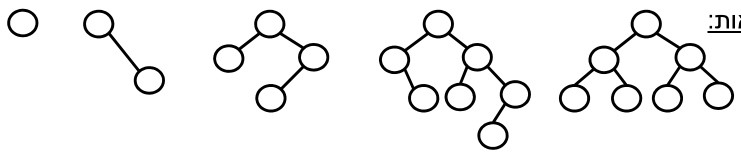
<u>הגדרה</u>: נסמן ב-$h(T)$ גובה של עץ T. משפחת עצים תקרא <u>מאוזנת</u> אם לכל עץ T בן n צמתים במשפחה מתקיים:

$h(T) = O(\log n)$.

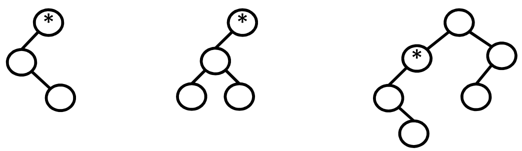
<u>**עצי AVL (Adelson-Velsky, Landis)**</u>

<u>הגדרה</u>: עץ AVL הוא עץ חיפוש בינרי שבו לכל צומת v התכונה:
$|h(v\to left) - h(v\to right)| \le 1$

<u>דוגמאות:</u>



<u>דוגמאות נגד:</u>



* צומת בו מופר האיזון

תרשים הוכחה למציאת חסם לגובה עץ AVL

1. נגדיר משפחת עצים - עצי פיבונצ'י (Fibonacci trees): $F_0, F_1, F_2, \ldots, F_h, \ldots$
2. נראה שמספר הצמתים בעץ F_h מקיים $|F_h| \ge a^h$ עבור קבוע a כלשהו.
3. נראה ש<u>לכל</u> עץ AVL, T, בעל $n = |T|$ צמתים מגובה h יש מספר צמתים גדול\שווה מ-$|F_h|$.

<u>מסקנה</u>: לכל עץ AVL עם n צמתים וגובה h מתקיים: $n \ge |F_h| \ge a^h$.

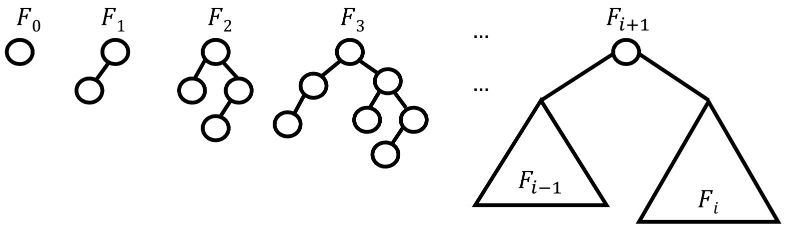
מכאן נקבל כי $\log_a n \ge h$.

<u>מסקנה</u>: משפחת עצי AVL מאוזנת.

<u>הערה</u>: עצי פיבונצ'י הם עצי AVL בהם הגובה גדל הכי מהר כפונקציה של n.

חסם לגובה עץ AVL

נגדיר משפחת עצים - עצי פיבונצ'י (Fibonacci trees):



<u>טענה 1</u>: לכל h, לעץ F_h גובה h.

<u>הוכחה – באינדוקציה על h:</u>

<u>בסיס:</u> נכון עבור F_0 ועבור F_1.

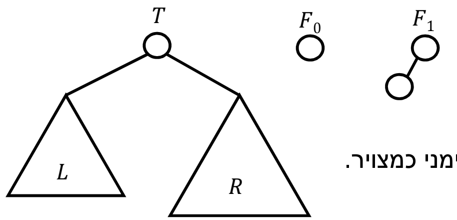
<u>צעד האינדוקציה:</u> מתקיים לפי הגדרת עץ פיבונצ'י:

$$height(F_{i+1}) = height(F_i) + 1 = i + 1$$

<u>טענה 2</u>: $|F_h| = 1 + |F_{h-1}| + |F_{h-2}| \ge |F_{h-1}|$.

חסם לגובה עץ AVL (המשך)

<u>טענה 3</u>: יהי T עץ AVL בעל גובה h. אזי $|T| \ge |F_h|$.



<u>הוכחה - באינדוקציה על h</u>:

<u>בסיס</u>: נכון עבור $h = 0$ ועבור $h = 1$:

<u>צעד האינדוקציה</u>:

יהי T עץ AVL בעל גובה h.

יהי L תת-העץ השמאלי ו-R תת-העץ הימני כמצויר.

- תתי העצים R ו- L הם עצי AVL בגובה קטן מהגובה של T.
- אחד מהם (R נאמר) בגובה $h-1$ והשני בגובה $h-1$ או $h-2$.
- לפי הנחת האינדוקציה, $|R| \ge |F_{h-1}|$.
- לפי הנחת האינדוקציה וטענה 2, $|L| \ge \min\{|F_{h-1}|, |F_{h-2}|\} = |F_{h-2}|$.
- לפיכך T הוא עץ בן לפחות $|F_h|$ צמתים שכן מתקיים:

$$|T| = 1 + |R| + |L| \ge 1 + |F_{h-1}| + |F_{h-2}| = |F_h|$$

מספרי פיבונצ'י

סדרת פיבונצ'י n_i מוגדרת באופן הבא :

$$n_0 = 0, \qquad n_1 = 1$$

כששאר איברי הסדרה מוגדרים לפי:

$$n_{i+1} = n_i + n_{i-1}$$

כמה איברים ראשונים בסדרה:

$n_0 = 0,\quad n_1 = 1,\quad n_2 = 1$
$n_3 = 2,\quad n_4 = 3,\quad n_5 = 5$
$n_6 = 8,\quad n_7 = 13,\quad n_8 = 21$

<u>טענה 4</u>: $n_i = \frac{\phi^i - \bar{\phi}^i}{\sqrt{5}}$, כאשר $\bar{\phi} = \frac{1-\sqrt{5}}{2}$ וכן $\phi = \frac{1+\sqrt{5}}{2}$ (ϕ נקרא יחס הזהב).

<u>הערה:</u> מכיוון שמתקיים $|\bar{\phi}| < 1 < |\phi|$ נובע גם כי $n_i \approx \frac{\phi^i}{\sqrt{5}}$ עבור i גדול.

מספרי פיבונצ'י (המשך)

<u>הוכחת טענה 4</u>:

נתונה המשוואה הבאה: $n_{i+1} = n_i + n_{i-1} \quad n_0 = 0 \quad n_1 = 1$

נניח פתרון מהצורה: $n_i = a \cdot x^i$

נציב במשוואה ונקבל: $a \cdot x^{i+1} = a \cdot x^i + a \cdot x^{i-1}$

נצמצם ב-$a \cdot x^{i-1}$ ונקבל $x^2 = x + 1$

<u>הבחנה 1</u>: פתרונות משוואה זו הם $\phi = \frac{1+\sqrt{5}}{2}$ ו- $\bar{\phi} = \frac{1-\sqrt{5}}{2}$.

משוואת ההפרשים ליניארית ולכן כל צרוף ליניארי של פתרונות מהווה פתרון:

$$n_i = a \cdot \phi^i + b \cdot \bar{\phi}^i$$

<u>שימוש בתנאי השפה מוביל למציאת הקבועים</u>:

$$n_0 = 0 \;\Rightarrow\; a \cdot \phi^0 + b \cdot \bar{\phi}^0 = a + b = 0 \;\Rightarrow\; b = -a$$

$$n_1 = 1 \;\Rightarrow\; a \cdot \left(\frac{1+\sqrt{5}}{2}\right) - a \cdot \left(\frac{1-\sqrt{5}}{2}\right) = 1 \Rightarrow \; a = \frac{1}{\sqrt{5}}$$

לפיכך פתרון המשוואה הוא: $n_i = \frac{\phi^i - \bar{\phi}^i}{\sqrt{5}}$.

<u>הערה:</u> כמו בהרבה מקרים, גם לטענה 4 מספר הוכחות. בחוברת מוצגות 2 הוכחות נוספות לטענה זו.

מספרי פיבונצ'י (המשך)

<u>טענה 4</u>: $n_i = \frac{\phi^i - \bar{\phi}^i}{\sqrt{5}}$, כאשר $\bar{\phi} = \frac{1-\sqrt{5}}{2}$ וכן $\phi = \frac{1+\sqrt{5}}{2}$ (ϕ נקרא יחס הזהב).

<u>הוכחה שניה לטענה 4 - באינדוקציה על i</u>:

<u>בסיס</u>: עבור $i = 0$ מתקיים: $n_0 = 0 = \frac{\phi^0 - \bar{\phi}^0}{\sqrt{5}}$

עבור $i = 1$ מתקיים: $n_1 = 1 = \frac{\phi^1 - \bar{\phi}^1}{\sqrt{5}} = \left(\frac{1+\sqrt{5}}{2} - \frac{1-\sqrt{5}}{2}\right) \cdot \frac{1}{\sqrt{5}}$

<u>צעד</u>: נניח כי הטענה נכונה עבור k ו- $k+1$.

נראה כי מכאן נובע עבור $i = k+2$.

מהגדרת מספרי פיבונצ'י + הנחת האינדוקציה:

$$n_{k+2} = n_{k+1} + n_k = \frac{\phi^{k+1} - \bar{\phi}^{k+1}}{\sqrt{5}} + \frac{\phi^k - \bar{\phi}^k}{\sqrt{5}}$$

$$= \frac{(\phi + 1) \cdot \phi^k - (\bar{\phi} + 1) \cdot \bar{\phi}^k}{\sqrt{5}}$$

מהבחנה 1 $\Rightarrow$ $$= \frac{(\phi^2) \cdot \phi^k - (\bar{\phi}^2) \cdot \bar{\phi}^k}{\sqrt{5}} = \frac{\phi^{k+2} - \bar{\phi}^{k+2}}{\sqrt{5}}$$

מספרי פיבונצ'י (המשך)

<u>הוכחת שלישית לטענה 4 – אלגברה לינארית:</u>

נתונה המשוואה הבאה: $n_1 = 1 \quad n_0 = 0 \quad n_{i+1} = n_i + n_{i-1}$

נתייחס לפעולת המטריצה Aהבאה:

$$\begin{pmatrix}1&1\\1&0\end{pmatrix}\begin{pmatrix}n_i\\n_{i-1}\end{pmatrix} = \begin{pmatrix}n_{i+1}\\n_i\end{pmatrix}$$

אם נכפול את המטריצה מספר פעמים, נקבל כי (ניתן להראות באינדוקציה):

$$\begin{pmatrix}1&1\\1&0\end{pmatrix}^i\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}1&1\\1&0\end{pmatrix}^i\begin{pmatrix}n_1\\n_0\end{pmatrix} = \begin{pmatrix}n_{i+1}\\n_i\end{pmatrix}$$

כדי לקבל נוסחה סגורה, נמצא תחילה את הערכים העצמיים (ע"ע) של A, ע"י מציאת שורשים לפולינום האופייני של אותה A:

$$p_A(x) = \det(A - x\cdot I) = det\begin{pmatrix}1-x & 1\\ 1 & -x\end{pmatrix} = x^2 - x - 1$$

אך אנו כבר יודעים מהם השורשים של משוואה זו: אלו $\phi = \frac{1+\sqrt5}{2}$ ו- $\bar\phi = \frac{1-\sqrt5}{2}$

עתה, נמצא מהם הווקטורים העצמיים (ו"ע) המתאימים לאותם הע"ע, נתאר את הווקטור $\binom{1}{0}$ בתור צירוף לינארי של אותם ו"ע ומכאן נקבל את הנוסחה הסגורה.

מספרי פיבונצ'י (המשך)

<u>הוכחת שלישית לטענה 4 – אלגברה לינארית (המשך):</u>

בשקף הקודם ראינו כי מתקיים:

$$A^i\cdot\binom{1}{0} = \begin{pmatrix}1&1\\1&0\end{pmatrix}^i\cdot\binom{1}{0} = \begin{pmatrix}1&1\\1&0\end{pmatrix}^i\cdot\binom{n_1}{n_0} = \binom{n_{i+1}}{n_i}$$

כמו-כן, מצאנו כי הע"ע של המטריצה A הם $\phi = \frac{1+\sqrt5}{2}$ ו- $\bar\phi = \frac{1-\sqrt5}{2}$

- נזכיר כי ווקטור עצמי של A המתאים לע"ע λ הוא ווקטור $v_\lambda = \binom{a}{b}$ המקיים $A\cdot v_\lambda = \lambda v_\lambda$. במקרה שלנו נדרוש $\begin{pmatrix}1&1\\1&0\end{pmatrix}\cdot\binom{a}{b} = \binom{a+b}{a} = \binom{\lambda a}{\lambda b}$.
- מכאן נקבל כי ו"ע של ϕ הוא $v_\phi = \binom{\phi}{1}$ וכי ו"ע של $\bar\phi$ הוא $v_{\bar\phi} = \binom{\bar\phi}{1}$.
- נבחין כי $\frac{1}{\sqrt5}(\phi - \bar\phi) = \frac{1}{\sqrt5}\left(\frac{1+\sqrt5}{2} - \frac{1-\sqrt5}{2}\right) = 1$, וכי $\frac{1}{\sqrt5}(1-1) = 0$.
- כלומר, $\binom{1}{0} = \frac{1}{\sqrt5}(v_\phi - v_{\bar\phi})$. נציב זאת בנוסחה מתחילת השקף ונקבל:

$$\binom{n_{i+1}}{n_i} = A^i\cdot\binom{1}{0} = \frac{1}{\sqrt5}A^i\cdot(v_\phi - v_{\bar\phi}) = \frac{1}{\sqrt5}\binom{\phi^{i+1}}{\phi^i} - \frac{1}{\sqrt5}\binom{\bar\phi^{i+1}}{\bar\phi^i}$$

ולפי הקואורדינטה השנייה בביטוי הנ"ל נקבל כי אכן: $n_i = \frac{\phi^i - \bar\phi^i}{\sqrt5}$.

ניתוח גובה עץ AVL

<u>טענה 5</u>: לעץ פיבונצ'י F_i יש $|F_i| = n_{i+3} - 1$ צמתים, כש- n_i הוא מספר פיבונצ'י ה-i .

<u>הוכחה</u>: מתקיימת המשוואה הרקורסיבית הבאה:

$$|F_0| = 1, |F_1| = 2, \qquad |F_{i+1}| = |F_i| + |F_{i-1}| + 1$$

במקום לפתור משוואה זו ישירות, נבצע שנוי משתנים כדי לקבל את המשוואה שכבר פתרנו.

יהי t_i מספר הצמתים ב-F_i ועוד אחד. כלומר מתקיים $|F_i| = t_i - 1$.

נציב במשוואה הרקורסיבית ונקבל: $t_{i+1} - 1 = (t_i - 1) + (t_{i-1} - 1) + 1$

כלומר, $t_{i+1} = t_i + t_{i-1}$

$t_0 = 2, \ \ t_1 = 3$

זו המשוואה של מספרי פיבונצ'י כשנקודת ההתחלה מוזזת בשלושה אינדקסים ולכן מתקיים: $|F_i| + 1 = t_i = n_{i+3}$.

הוכחה פורמלית מתקבלת באינדוקציה על i:

<u>בסיס:</u> $t_0 = 2 = n_3, \quad t_1 = 3 = n_4$

<u>צעד האינדוקציה:</u> $t_{i+1} = t_i + t_{i-1} = n_{i+3} + n_{i+2} = n_{i+4}$

ניתוח גובה עץ AVL (המשך)

<u>טענה 6</u>: יהי T עץ AVL בן n צמתים וגובה h, אזי $h = O(\log n)$.

- לאור טענה 3 מתקיים:

$$n = |T| \ge |F_h|$$

- לאור טענות 5, 4 מתקיים:

$$n \ge |F_h| = n_{h+3} - 1 = \frac{\phi^{h+3} - \bar\phi^{h+3}}{\sqrt5} - 1 \underset{\boxed{1 > \bar\phi^{h+3}}}{\ge} \frac{\phi^{h+3}}{\sqrt5} - 2$$

- לקיחת לוגריתם משני צדי המשוואה:

$$\log_\phi\left(\sqrt5(n+2)\right) \ge h + 3$$

- נבודד את h ונקבל:

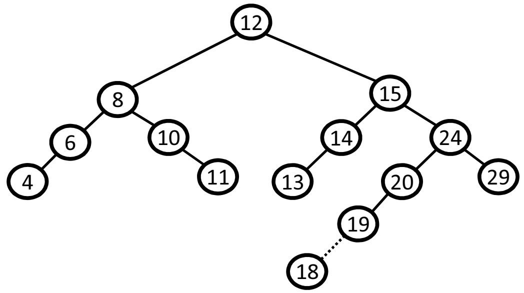
$$\log_\phi(n+2) + \log_\phi\sqrt5 - 3 \ge h$$

- לסיכום:

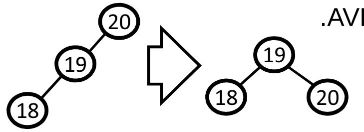
$$h = O(\log n)$$

איזון בעץ AVL

מטענה 6 נובע שזמן החיפוש בעץ AVL הוא $O(\log n)$. נצטרך לדאוג שלאחר הכנסה או הוצאה, העץ הנותר יהיה עץ AVL.



לאחר הוספת האיבר 18 מתקבל עץ שאינו עץ AVL.
אבל ניתן לשנות את תת העץ שבו הופר האיזון בצורה הבאה:
תיקון כזה נקרא <u>גלגול</u>.



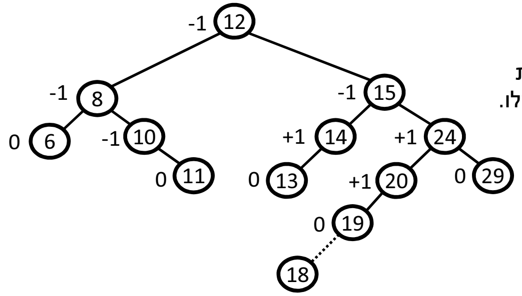
בזמן הוצאה תיתכן הפרת איזון דומה. למשל, בעץ הנ"ל, בהוצאת 29.

איזון בעץ AVL (המשך)

<u>לכל צומת v בעץ בינרי נסמן:</u>
$h_L(v)$ - גובה תת העץ השמאלי של v.
$h_{R(v)}$ - גובה תת העץ הימני של v.
גורם האיזון (Balance Factor) מחושב כהפרש הגבהים:
$$BF(v) = h_L(v) - h_R(v)$$

<u>לדוגמא:</u>

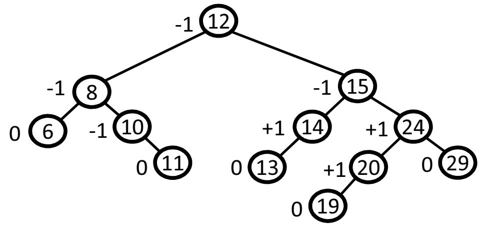
- מצד שמאל לכל צומת מסומן גורם האיזון שלו.



אחרי ההכנסה של 18 גורם האיזון מופר על מסלול ההכנסה.

הבחנות

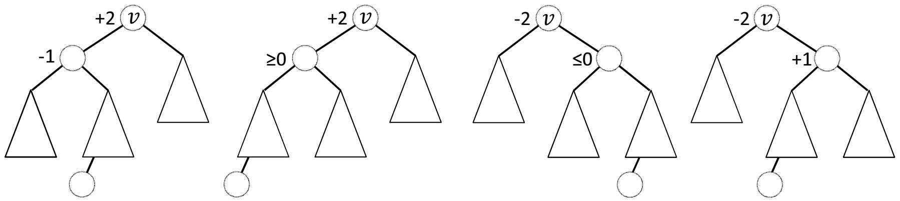
1. הצמתים היחידים שאולי הופר בהם האיזון הם הצמתים לאורך מסלול הכנסה/הוצאה.
2. אם עבור צומת v במסלול הנ"ל גובה העץ ששורשו v לא השתנה אזי גורמי האיזון בצמתים שמעליו במסלול לא השתנו.
3. אם גורם האיזון הופך ל-2 או ל 2-, אזי יש לבצע גלגול על מנת שהעץ יחזור להיות עץ AVL.
4. <u>גלגול</u> – פעולה המתבצעת על צומת שהופר בו האיזון על מנת להחזירו לתחום המותר [-1,0,1].
5. גורם האיזון לא יכול להיות גדול מ-2 בערכו המוחלט כי בכל הכנסה\הוצאה\גלגול הוא משתנה ב-1 לכל היותר.



סוגי הגלגולים

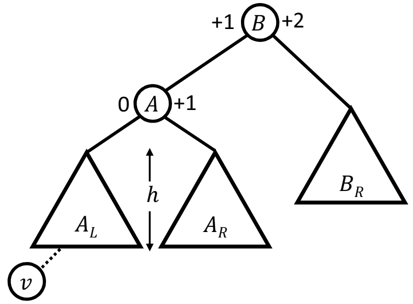
סוג הגלגול, כלומר הדרך לתקן חוסר איזון בצומת, תלוי בצורה בה האיזון מופר. נתן לסווג חוסר איזון בארבע קטגוריות שונות המכסות את כל המקרים:

גלגול RL גלגול RR גלגול LL גלגול LR



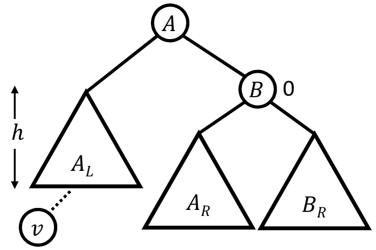
בשורש v	בבן השמאלי v_L	בבן השמאלי v_R	הגלגול המתאים
$BF(v) = 2$	$BF(v_L) \geq 0$		LL
$BF(v) = 2$	$BF(v_L) = -1$		LR
$BF(v) = -2$		$BF(v_R) \leq 0$	RR
$BF(v) = -2$		$BF(v_R) = 1$	RL

גלגול LL



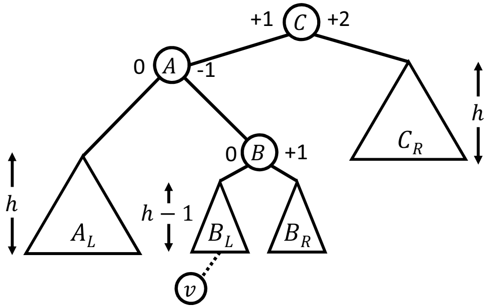
לפני הכנסת v: גובה העץ הוא $h+2$.
הוכנס צומת v שהגדיל את גובה A_L ל- $h+1$.
גלגול LL: יעביר את A לשורש.

- מצד ימין של הצמתים מסומנים גורמי האיזון אחרי ההכנסה.

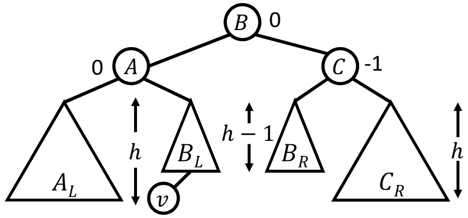


אחרי הכנסת v:
גובה העץ לאחר הגלגול הוא $h+2$, כמו לפני ההכנסה. כל הצמתים בתת העץ מאוזנים. שינינו $O(1)$ מצביעים ולכן זמן הגלגול $O(1)$.

גלגול LR



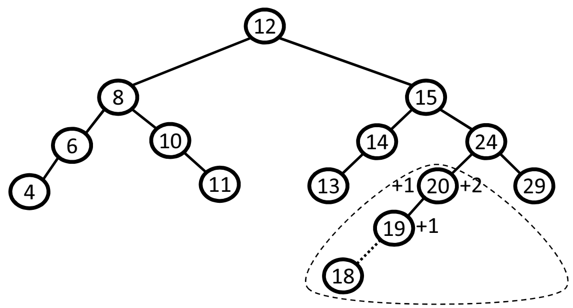
לפני הכנסת v: גובה העץ הוא $h+2$.
הוכנס צומת v שהגדיל את גובה B_L ל-h.
גלגול LR: יעביר את B לשורש.



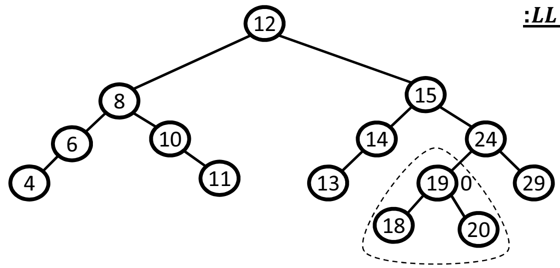
אחרי הכנסת v:
גובה העץ לאחר הגלגול הוא $h+2$, כמו לפני ההכנסה. כל הצמתים בתת העץ מאוזנים. שינינו $O(1)$ מצביעים ולכן זמן הגלגול $O(1)$.

דוגמא להכנסת ערך x לעץ AVL

הוסף 18:

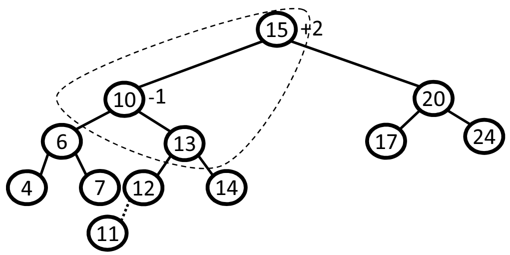


לאחר גלגול LL:

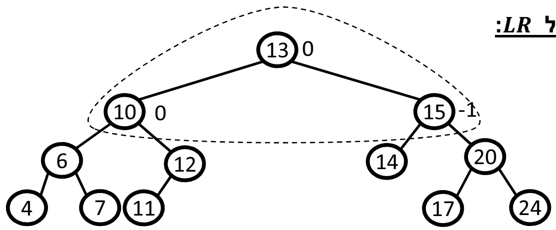


דוגמא להכנסת ערך x לעץ AVL

הוסף 11:

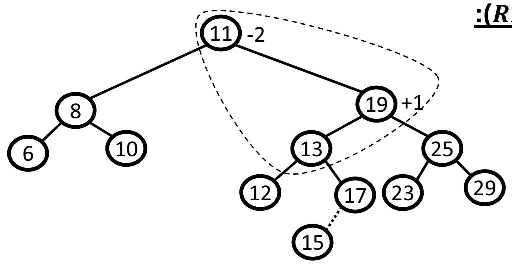


לאחר גלגול LR:

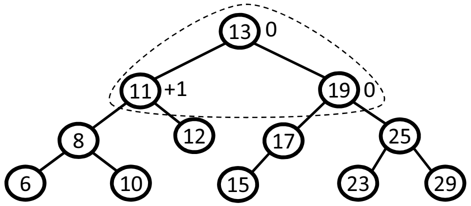


דוגמא להכנסת ערך x לעץ AVL

הוסף 15 (נחוץ גלגול RL):



לאחר גלגול RL:

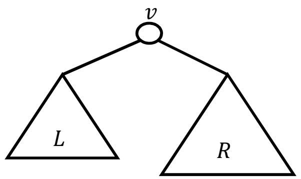


בהכנסה, לאחר גלגול אחד העץ מאוזן מכיוון שגובה תת העץ בו נעשה השנוי לא השתנה.

מימוש עץ AVL – שמירת שדה גובה

בכדי לחשב את $BF(v)$ בקלות, די לנו לשמור בכל צומת v שדה גובה, אותו נסמן ב-$h(v)$. $BF(v)$ הוא פשוט $h(v \to left) - h(v \to right)$.

חלק גדול מהנוחות של שדה זה היא קלות התחזוק שלו: השדה $h(v)$ ניתן לחישוב ב-$O(1)$ בהינתן השדות העדכניים $h(v \to left)$ ו- $h(v \to right)$:



$$h(v) = 1 + \max\{h(v \to left), h(v \to right)\}$$

למעשה, אם רק אחד השדות $h(v \to left)$ או $h(v \to right)$ השתנו (עקב הכנסה\הוצאה או גלגולים), מספיק לעדכן את $h(v)$ לפי תת העץ שגובהו השתנה.

אלגוריתם להכנסת ערך x לעץ AVL

1. הכנס את x כמו לעץ חיפוש בינרי. יהי v העלה שהוסף.
2. $h(v) \leftarrow 0$.
3. <u>כל עוד $v \neq root$ בצע:</u>
   1. $p \leftarrow parent(v)$.
   2. אם $h(p) \geq h(v) + 1$ סיים.
   3. $h(p) \leftarrow h(v) + 1$.
   4. אם ב- p הופר האיזון, בצע גלגול וסיים.
   5. אחרת $v \leftarrow p$.

<u>שאלה:</u> איך נחשב את $parent(v)$?

נוציא אותו מהמחסנית בה נמצאים כל הצמתים על המסלול מהשורש ועד v.

זמן ההכנסה לעץ AVL

כיוון שהצומת בו עושים גלגול חוזר לגובהו תרם ההכנסה, מבצעים רק גלגול אחד.

לכן, נקבל כי הזמן הנדרש להכנסת איבר לעץ AVL הוא:

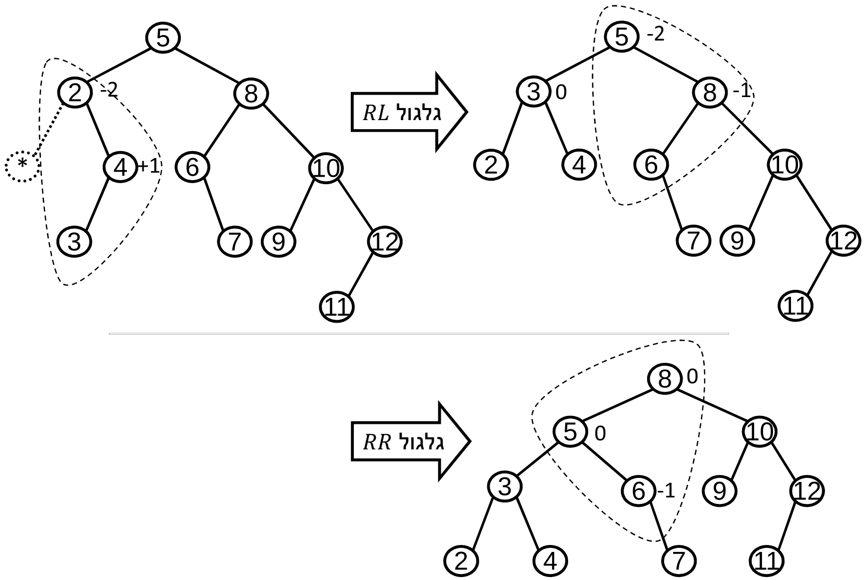
מציאת המקום הדרוש להכנסה	$O(h)$
הוספת הצומת	$O(1)$
מציאת המקום בו מופר האיזון (אם מופר)	$O(h)$
תיקון האיזון	$O(1)$
סה"כ	$O(h) = O(\log n)$

אלגוריתם הוצאה/הכנסה

1. הוצא (הכנס) צומת v כפי שהפעולה מתבצעת בעץ חיפוש בינרי.
2. תקן את גורמי האיזון בצורה הבאה: לכל צומת v לאורך המסלול החל מלמטה ועד לשורש בצע:
    1. עדכן את $h(v)$.
    2. אם $|BF(v)| = 2$, בצע גלגול והמשך כלפי מעלה.
    3. אם גובה תת העץ ששורשו v לא השתנה, סיים.
    4. אם גובה תת העץ השתנה ו- $BF(v)$ תקין, המשך כלפי מעלה.

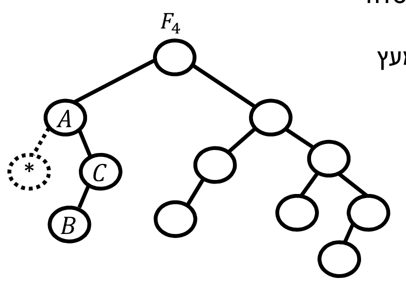
<u>הערה:</u> בהוצאה יתכן יותר מגלגול אחד.

דוגמא להוצאה מעץ AVL



דוגמא: הוצאה מעץ פיבונצ'י

בהוצאה יתכן גלגול בכל צומת על המסלול בין צומת המוצא ובין השורש.
<u>לדוגמא:</u> כאשר מוציאים עלה ראשון מעץ פיבונצ'י - F_i.



<u>תרגיל לבית:</u> מהי סדרת הגלגולים המתבצעת כאשר מוציאים עלה ראשון מעץ פיבונצ'י - F_i?

זמן ההוצאה מעץ AVL

מציאת המקום הדרוש להוצאה	$O(h)$
מציאת המקום בו מופר האיזון (אם מופר)	$O(h)$
תיקון האיזון (לכל היותר פעם בכל רמה)	$O(h)$
סה"כ	$O(h) = O(\log n)$

<u>לסיכום:</u> עצי AVL מאפשרים חיפוש, הכנסה, הוצאה בזמן $O(\log n)$.

ראו הדגמה באתר הקורס (בחרו links/AVL tree animation):
http://webcourse.cs.technion.ac.il/234218