Arbori binari - a.b. stricti şi a.b. echilibraţi AVL Aplicaţii la analiza performanţei căutării binare

Reamintim următoarele proprietaţi pentru un **arbore binar strict** T:

$N_E(T)$ (sau N_E) = numărul de frunze ale lui T;
$N_I(T)$ (sau N_I) = numărul de noduri interne ale lui T;
$L_E(T)$ (sau L_E) = lungimea externă a lui T;
$L_I(T)$ (sau L_I) = lungimea internă a lui T;

Propoziţia 1. În oricare a.b.s. T avem relaţia:

$$N_E = N_I + 1.$$

Propoziţia 2. În oricare a.b.s. T avem relaţia:

$$L_E = L_I + 2N_I.$$

Propoziţia 3. În oricare a.b.s. T de înălţime $h(T) = d$ avem relaţia:

$$N_E \leq 2^d.$$

Corolarul 1. În oricare a.b.s. T de înălţime $h(T) = d$ avem relaţia:

$$d \geq \lceil log_2 N_E \rceil.$$

Propoziţia 4. În familia a.b. stricţi cu număr de frunze fixat, N_E, lungimea externă minimă se atinge pentru aceia care au frunzele repartizate pe cel mult două niveluri adiacente.

Propoziţia 5. Lungimea externă minima a unui a.b.s. cu N_E frunze este:

$$L_E^{\min} = N_E \lfloor log_2 N_E \rfloor + 2(N_E - 2^{\lfloor log_2 N_E \rfloor}).$$

Din *Demonstratie:* dacă $d = h(T)$= înălţimea unui a.b.s. T pe care se atinge lungimea externă minima, avem 2 cazuri (cf. Prop. 4): Cazul (a): Toate frunzele sunt la acelaşi nivel, $d = h(T)$, daca $N_E = 2^d$. Cazul (b): Frunzele nu sunt toate la acelaşi nivel. Dar atunci ele sunt repartizate doar pe nivelurile $d-1$ (fie y nr. de frunze de la acest nivel), şi d (fie $2x$ nr. de frunze de la acest nivel, x= nr. de noduri interne de la nivelul $d-1$). Se rezolvă sistemul:

$$\begin{cases} x + y = 2^{d-1} & (1) \\ x + y = N_E & (2) \end{cases}$$

Avem:

$$\begin{cases} \text{nr. de frunze la nivelul } d-1 = y = 2^d - N_E, \\ \text{nr. de frunze la nivelul } d = 2x = 2N_E - 2^d. \end{cases}$$

Propoziţia 6. Intr-un a.b.s. *lungimea externă medie* $L_E^{\text{medie}} = \dfrac{L_E}{N_E}$ are marginea inferioară

$$L_E^{\text{medie}} \geq \lfloor log_2 N_E \rfloor.$$

Teorema AVL
Margine superioară şi margine inferioară pentru înălţimea unui arbore binar echilibrat AVL

Teorema 1 (***Teorema AVL***)**.** Fie T un arbore binar strict şi echilibrat AVL, cu n noduri interne. Fie $h(T)$ înălţimea lui. Avem:

$$log_2(n+1) \leq h(T) \leq 1.4404\ log_2(n+2) - 0.328.$$

Echivalent. Sunt satisfăcute următoarele inegalităţi:

(1) $h(T) \geq log_2(n+1)$.

(2) $h(T) \leq (1/log_2\phi)log_2(n+2) + (log_2 5/2log_2\phi - 2)$, unde $\phi = (1+\sqrt{5})/2$.

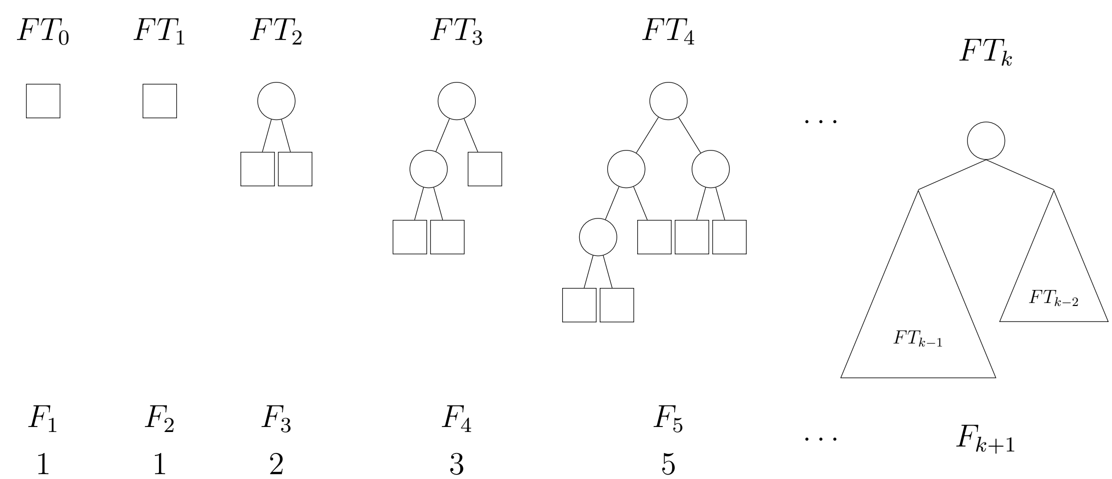
Demonstraţie: Inegalitatea (1) este adevarată pentru a.b.s. în general (rezultă din Corolarul la Prop. 3).

Pentru a **dem. ineg. (2)** construim o clasă particulară de a.b.s. şi echil. AVL, *arborii Fibonacci.*

Numerele Fibonacci (de ordinul 1): $F_1 = F_2 = 1$ şi relaţia de recurenţă $F_{n+2} = F_{n+1} + F_n$, pentru $n \geq 1$.

Formula Binet pentru numere Fibonacci: $F_n = (1/\sqrt{5})(\phi^n - \bar{\phi}^n)$, unde $\phi = (1+\sqrt{5})/2$.

Construim prin recurenţă familia de arbori binari $(FT_k)_{k\geq 0}$, FT_k = Arbore Fibonacci (Fib Tree) de ordin k.



Lema 1. Pentru orice $k \geq 0$ arborele FT_k este a.b.s.

Lema 2. Pentru orice $k \geq 1$ arborele FT_k are caracteristicile:

(a) $h(FT_k) = k - 1$.

(b) $N_E(FT_k) = F_{k+1}$.

(c) $N_I(FT_k) = F_{k+1} - 1$.

Dem: Este suficient să demonstrăm (a) şi (b), (c) este consecinţă a lui (b) prin Prop 1.
Inducţie după k, $k \geq 1$.
$k = 1$: FT_1 este ... cu $h(FT_1) = 0$ şi $N_E(FT_1) = 1 = F_2$.
Pp. (a) şi (b) adev. pentru FT_m, orice $m < k$.
Fie k oarecare, fixat, $k \geq 3$. Avem:

(a) $h(FT_k) = h(FT_{k-1}) + 1 = (k - 2) + 1 = k - 1$.

(b) $N_E(FT_k) = N_E(FT_{k-1}) + N_E(FT_{k-2}) = F_k + F_{k-1} = F_{k+1}$.

Lema 3. Pentru orice $k \geq 0$ arborele FT_k este echilibrat AVL.

Dem: Pt. $k = 0, 1, 2$ direct. Pt. $k \geq 3$, (inductie), de dem. in nodul radacina se fol. (a) din Lema 2.

Lema 4. În familia arborilor binari stricţi şi echilibraţi AVL de înălţime data, h, arborii Fibonacci au număr minim de noduri interne.

Dem: Inducţie după h.
$h = 0$. Singurii a.b.Fib. ... au $N_I = 0$.
$h = 1$. T_1= a.b.s. de înălţime 1 şi nr minim de noduri interne, are 1 nod intern (rădăcina) şi 2 frunze, i.e. $T_1 = FT_2$.
Notăm cu T_h un a.b.s. şi echil. AVL de înălţime h care are nr. minim de noduri interne.
Ipot. inducţie: pentru orice k, $k < h$ avem $T_k = FT_{k+1}$.
h oarecare, $h \geq 2$: fie T_h ca mai sus. Are nod rad. cu fii $left(T_h)$ şi $right(T_h)$. Putem pp. ca $h(left(T_h)) > h(right(T_h))$. Avem:

(i) $h(left(T_h)) = h - 1$ şi $N_I(left(T_h))$ minim, deci $left(T_h) = T_{h-1}$.

(ii) $h(right(T_h)) = h - 2$ şi $N_I(right(T_h))$ minim, deci $right(T_h) = T_{h-2}$.

Dar, cf. ipot. ind., $T_{h-1} = FT_h$ şi $T_{h-2} = FT_{h-1}$, deci, din (i) $left(T_h) = FT_h$ si din (ii) $right(T_h) = FT_{h-1}$, din care rezulta ca $T_h = FT_{h+1}$.
Observatie: Cf. Lemei 1 nr. minim de noduri interne pentru înălţime h dată va fi $N_I(FT_{h+1}) = F_{h+2} - 1$.

Revenim la dem. Th. ineg. (2).

Fie T un a.b.s. şi echil AVL, cu $n = N_I(T)$ noduri interne şi înălţime $h = h(T)$. Cf. Obs. de după lema 4, avem

$$n \geq F_{h+2} - 1.$$

$$\frac{1}{\sqrt{5}}\phi^{h+2} - \frac{1}{\sqrt{5}}\bar{\phi}^{h+2} - 1 \leq n,$$

unde $\phi = \frac{1+\sqrt{5}}{2}$, $\bar{\phi} = \frac{1-\sqrt{5}}{2}$. Din $-1 \leq \bar{\phi} \leq 0$ rezultă

$$n \geq \frac{1}{\sqrt{5}}\phi^{h+2} - 2,$$

$$n + 2 \geq \frac{1}{\sqrt{5}}\phi^{h+2},$$

$$log_2(n+2) \geq (h+2)log_2\phi - \frac{1}{2}log_2 5.$$

Desfac, în fcţ. de h ... rezultă

$$h \leq \frac{1}{log_2\phi}log_2(n+2) + \frac{log_2 5}{2log_2\phi} - 2 = a\ log_2(n+2) + b,$$

si $a < 1.4404$, $b < -0.328$.$\blacksquare$