ALS1 – Přednáška 3

1 Catalanova čísla

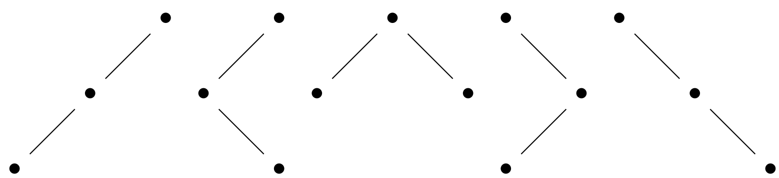
Počet stromů o 0 uzlech: 1

$$\emptyset$$

Počet stromů o 1 uzlu: 1

Počet stromů o 2 uzlech: 2

Počet stromů o 3 uzlech: 5



Počet stromů o 5 uzlech: 14

1.1 Kolik je binárních vyhledávacích stromů?

$$\begin{aligned}
C_1 &= C_0C_0\\
C_2 &= C_1C_0 + C_0C_1\\
C_3 &= C_2C_0 + C_1C_1 + C_0C_2\\
C_4 &= C_3C_0 + C_2C_1 + C_1C_2 + C_0C_3\\
&\vdots\\
C_n &= C_{n-1}C_0 + C_{n-2}C_1 + \cdots + C_1C_{n-2} + C_0C_{n-1}
\end{aligned}$$

1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, 58786, 208012, 742900, 2674440, 9694845, 35357670, 129644790, 477638700, 1767263190, 6564120420, 24466267020, 91482563640, 343059613650, 1289904147324, ...

tzv. Catalanova čísla

1.2 Explicitní vzorec pro Catalanova čísla

Definujeme *generující polynom*

$$f(z) = C_0 + C_1z + C_2z^2 + C_3z^3 + \cdots = \sum_{i=0}^{\infty} C_iz^i$$

obsahující všechna Catalanova čísla jako koeficienty.

uvažujme druhou mocninu $f(z)$:

$$\begin{aligned}
[f(z)]^2 = \;& C_0C_0 && (= C_1)\\
&+ (C_1C_0 + C_0C_1)z && (= C_2z)\\
&+ (C_2C_0 + C_1C_1 + C_0C_2)z^2 && (= C_3z^2)\\
&+ \cdots
\end{aligned}$$

Tedy

$$[f(z)]^2 = C_1 + C_2 z + C_3 z^2 + C_4 z^3 + \cdots$$

Když předchozí rovnici vynásobíme z a přičteme C_0:

$$f(z) = C_0 + z[f(z)]^2$$

To je pouze kvadratická rovnice v $f(z)$, můžeme ji řešit známým způsobem:

$$zf^2 - f + C_0 = 0.$$

Dostaneme

$$f(z) = \frac{1 \pm \sqrt{1-4z}}{2z} \tag{1}$$

Používejeme jenom $-$ namísto $\pm$, protože víme, že $f(0) = C_0 = 1$ (kdybychom použili $+$, pak pro $z \to 0$ je $f(z) \to \infty$). Pro vyjádření $f(z)$ použijeme (zobecněný) binomický rozklad na

$$\sqrt{1-4z} = (1-4z)^{1/2}$$

$$(a+b)^n = \sum_{k=0}^{\infty} \binom{n}{k} a^{n-k} b^k = a^n + \frac{n}{1} a^{n-1} b + \frac{n(n-1)}{2 \cdot 1} a^{n-2} b^2 + \frac{n(n-1)(n-2)}{3 \cdot 2 \cdot 1} a^{n-3} b^3 + \cdots$$

Dostáváme

$$(1-4z)^{1/2} = 1 - \frac{\frac{1}{2}}{1} 4z + \frac{(\frac{1}{2})(-\frac{1}{2})}{2 \cdot 1}(4z)^2 - \frac{(\frac{1}{2})(-\frac{1}{2})(-\frac{3}{2})}{3 \cdot 2 \cdot 1}(4z)^3 + \frac{(\frac{1}{2})(-\frac{1}{2})(-\frac{3}{2})(-\frac{5}{2})}{4 \cdot 3 \cdot 2 \cdot 1}(4z)^4 - \frac{(\frac{1}{2})(-\frac{1}{2})(-\frac{3}{2})(-\frac{5}{2})(-\frac{7}{2})}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}(4z)^5 - \cdots$$

Což je

$$(1-4z)^{1/2} = 1 - \frac{1}{1!} 2z - \frac{1}{2!} 4z^2 - \frac{3 \cdot 1}{3!} 8z^3 - \frac{5 \cdot 3 \cdot 1}{4!} 16z^4 - \frac{7 \cdot 5 \cdot 3 \cdot 1}{5!} 32z^5 - \cdots$$

Dosadíme do (1) a dostaneme

$$f(z) = 1 + \frac{1}{2!} 2z + \frac{3 \cdot 1}{3!} 4z^2 + \frac{5 \cdot 3 \cdot 1}{4!} 8z^3 + \frac{7 \cdot 5 \cdot 3 \cdot 1}{5!} 16z^4 + \cdots$$

Nepohodlných součinů typu $7 \cdot 5 \cdot 3 \cdot 1$ se zbavíme následovně.

Všimněme si, že:

$$\begin{aligned} 2^2 \cdot 2! &= 4 \cdot 2 \\ 2^3 \cdot 3! &= 6 \cdot 4 \cdot 2 \\ 2^4 \cdot 4! &= 8 \cdot 6 \cdot 4 \cdot 2. \end{aligned}$$

atd. Tedy

$$\begin{aligned} f(z) &= 1 + \frac{1}{2}\left(\frac{2!}{1!1!}\right) z + \frac{1}{3}\left(\frac{4!}{2!2!}\right) z^2 + \frac{1}{4}\left(\frac{6!}{3!3!}\right) z^3 + \frac{1}{5}\left(\frac{8!}{4!4!}\right) z^4 \cdots \\ &= \sum_{i=0}^{\infty} \frac{1}{i+1} \binom{2i}{i} z^i. \end{aligned}$$

Z tohoto

$$f(z) = \sum_{i=0}^{\infty} \frac{1}{i+1} \binom{2i}{i} z^i$$

tedy dostáváme explicitní vzorec pro Catalanova čísla:

$$C_i = \frac{1}{i+1} \binom{2i}{i}$$

Binárních vyhledávacích stromů o i uzlech je tedy $\frac{1}{i+1}\binom{2i}{i}$.

Catalanův trojúhleník

1						
1	1					
1	2	2				
1	3	5	5			
1	4	9	14	14		
1	5	14	28	42	42	
1	6	20	48	90	132	132

2 Vyvážené stromy

Theorem 1. *Výška AVL-stromu s n vnitřními uzly je $\mathcal{O}(\log_2 n)$.*

Důkaz. Jaký je minimální počet uzlů v AVL-stromu výšky h?

Nechť T_h je takový strom nejmenším počtem uzlů; pak

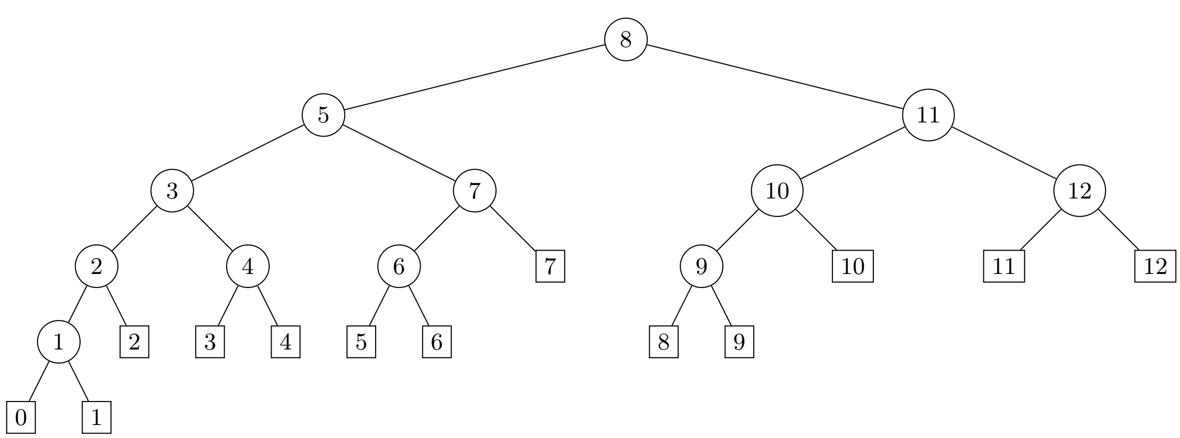
- (řekněme) levý podstrom má výšku $h-1$,
- pravý podstrom má výšku $h-1$ nebo $h-2$.

Protože chceme minimum uzlů, můžeme předpokládát, že levý podstrom kořene je T_{h-1} a pravý je T_{h-2}.

Intermezzo: Fibonacciho stromy

Fibonachiho strom řádu k má $F_{k+1}-1$ interních uzlů (na obrázku kroužky) a F_{k+1} externích uzlů (na obrázku čtverce), a je konstruován následovně:

- Pokud $k=0$ nebo $k=1$, strom je $\boxed{0}$.
- Pokud $k \geq 2$ kořen je F_k; levý podstrom je Fibonacciho strom řádu $k-1$; a pravý podstrom je Fibonacciho strom řádu $k-2$ se všemi čísly zvýšenými o F_k.



Obrázek 1. Fibonacciho strom řádu 8.

Vyjma externích uzlů se čísla obou potomků potomků v každém interním uzlu liší od čísla jejich předka o stejnou hodnotu, a tato hodnota je Fibonacciho číslo (např.: $5 = 8 - F_4$ a $11 = 8 + F_4$). Pokud je rozdíl F_j, odpovídající rozdíl pro další větev nalevo je F_{j-1}, zatímco napravo je to F_{j-2}.

Intermezzo: Binetův vzorec

Vyjádření Fibonachiho čísla:

$$F_n = \frac{\varphi^n - \psi^n}{\varphi - \psi} = \frac{\varphi^n - \psi^n}{\sqrt{5}},$$

kde

$$\varphi = \frac{1+\sqrt{5}}{2} \approx 1.6180339887\cdots \qquad \psi = \frac{1-\sqrt{5}}{2} \approx -0.6180339887\cdots$$

Čísla φ a ψ jsou řešení rovnic

$$x^2 = x + 1 \quad \text{a} \quad x^n = x^{n-1} + x^{n-2},$$

takže mocniny φ a ψ splňují Fibonacciho rekurenci. To jest,

$$\varphi^n = \varphi^{n-1} + \varphi^{n-2} \quad \text{a} \quad \psi^n = \psi^{n-1} + \psi^{n-2}.$$

Zjevně pro jakékoli hodnoty a, b, sekvence definovaná jako

$$U_n = a\varphi^n + b\psi^n \tag{$*$}$$

také splňuje stejnou rekurenci:

$$U_n = a\varphi^{n-1} + b\psi^{n-1} + a\varphi^{n-2} + b\psi^{n-2} = U_{n-1} + U_{n-2}$$

Pokud a, b určíme tak, aby $U_0 = 0, U_1 = 1$. Výsledná posloupnost musí být Fibonacciho. Takže potřebujeme vyřešit tuto soustavu rovnic:

$$\begin{aligned} a + b &= 0 \\ \varphi a + \psi b &= 1 \end{aligned}$$

Řešení je $a = \frac{1}{\varphi - \psi} = \frac{1}{\sqrt{5}}, b = -a$, to se dosadí do $(*)$ a hotovo.

konec intermezza

Máme tedy, že levý podstrom kořene je T_{h-1} a pravý je T_{h-2}.

Tedy Fibonachiho strom řádu $h+1$ má nejmenší počet uzlů mezi všemi vyváženými stromy výšky h.
Když vezmeme v úvahu, že $\psi^{h+2}/\sqrt{5} < 1$, dostáváme.

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

Zlogaritmujeme a hotovo.

Mohli bychom taky ukázat, že výška AVL-stromu o n prvcích je mezi

$$\log(n+1) \qquad \text{a} \qquad 1.4404lg(n+2) - 0.3277.$$

Analýza vkládání do AVL stromu: horní hranici času *v nejhorším případě* můžeme snadno odvodit z Věty o výšce AVL-stromu. Pro průměrný případ žádná komplexní analýza neexistuje. Na to je ten algoritmus už příliš složitý. Pořád je tu ale dost zajímavých teoretických a empirických výsledků:

2.1 Počet AVL stromů

Jaký je počet B_{nh} vyvážených binárních stromů s n vnitřními uzly a výškou h?

Pro malá h ze vztahů

$$\begin{aligned} B_0(z) &= 1 \\ B_1(z) &= z \\ B_{h+1}(z) &= zB_h(z)(B_h(z) + 2B_{h-1}(z)) \end{aligned}$$

Generující posloupnost

$$B_h(z) = \sum_{b \geq 0} B_{nh} z^n$$

Tedy např:

$$\begin{aligned} B_2(z) &= 2z^2 + z^3 \\ B_3(z) &= 4z^4 + 6z^5 + 4z^6 + z^7 \\ B_4(z) &= 16z^7 + 32z^8 + 44z^9 + \cdots + 8z^{14} + z^{15} \end{aligned}$$

Celkový počet vyvážených stromů s výškou h je $B_h = B_h(1)$, které splňuje rekurenci:

$$\begin{aligned} B_0 &= 1 \\ B_1 &= 1 \\ B_{h+1} &= B_h^2 + 2B_h B_{h-1} \end{aligned}$$

Takže

$$\begin{aligned} B_2 &= 1^2 + 2 \cdot 1 \cdot 1 = 3 && = 3 \\ B_3 &= 3^2 + 2 \cdot 3 \cdot 1 = 15 && = 3 \cdot 5 \\ B_4 &= 15^2 + 2 \cdot 15 \cdot 3 = 315 && = 3^2 \cdot 5 \cdot 7 \\ B_5 &= \ldots && = 3^3 \cdot 5^2 \cdot 7 \cdot 23 \end{aligned}$$

Obecně máme

$$B_h = A_0^{F_h} A_1^{F_{h-1}} \cdots A_{h-1}^{F_1} \cdot A_h^{F_0},$$

kde $A_0 = 1, A_1 = 3, A_2 = 5, A_3 = 7, A_4 = 23, A_5 = 347, \ldots, A_h = A_{h-1}B_{h-2} + 2$.

Posloupnosti B_h a A_h rostou velmi rychle – jsou *dvojitě exponenciální*. Lze ukázat, že existuje reálné číslo $\theta \approx 1.43687$ t.ž.

$$B_h = \lfloor \theta^{2^h} \rfloor - \lfloor \theta^{2^{h-1}} \rfloor + \lfloor \theta^{2^{h-2}} \rfloor - \cdots + (-1)^h \lfloor \theta^{2^0} \rfloor$$

Pokud uvažujeme B_h stromy stejně pravděpodobné, je průměrný počet uzlů ve stromu výšky h

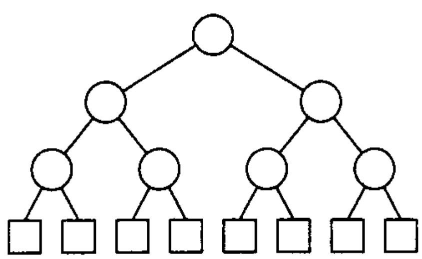
$$B_h'(1)/B_h(1) \approx (0.70118)2^h - 1. \tag{x}$$

To indikuje, že výška vyváženého stromu s n uzly bude obvykle daleko blíže $\log_2 n$ než k $\log_\theta n$.

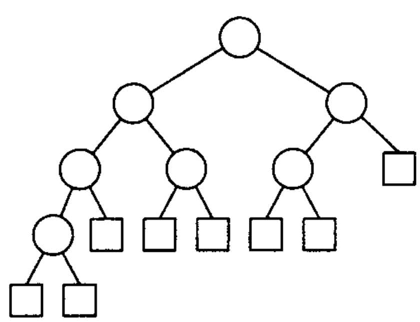
Naneštěstí tento výsledek nemá moc společného s algoritmem vkládání do AVL-stromu. Při jeho použití vznikají některé stromy se značně vyšší pravděpodobností, než jiné:

Např. uvažme případ $n = 7$, kde existuje 17 možných AVL-stromů, a $7! = 5040$ možných pořadí, ve kterých zle vložit 7 klíčů.

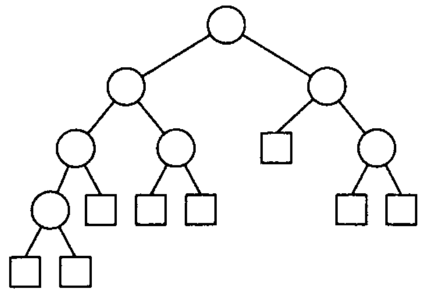
Perfektně vyvážený strom obdržíme v 2160 případech:



Fibonachiho strom obdržíme v 144 případech:



podobný strom obdržíme v 216 případech:



Fakt, že perfektně vyvážený strom vznikne vysokou pravděpodobností společně se vztahem (x).
Z toho bychom mohli usuzovat, že průměrně bude pro hledání v AVL potřeba okolo $\log_2 n + c$ porovnání.

Ovšem koeficient u termu $\log_2 n$ nebude přesně 1, protože kořen stromu by pak byl blízko mediánu a kořeny jeho podstromu blízko kvartilům. Jednoduchá nebo dvojitá rotace be nemohla jednoduše udržet kořen blízku mediánu.

Empirické testy ukazují, že průměrný počet porovnání pro vložení n-tého prvku je asi $1.01 \log_2 n + 0.1$, vyjma případů, kdy n je malé.