

ABZÄHLPROBLEME BEI BÄUMEN

HELMUT PRODINGER(WIEN)

Es sei \mathcal{B} die Familie der binären Bäume. \mathcal{B} kann durch eine formale Gleichung beschrieben werden:

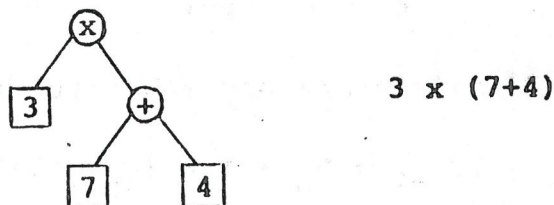
$$\mathcal{B} = \square + \begin{array}{c} \circ \\ / \quad \backslash \\ \mathcal{B} \quad \mathcal{B} \end{array}$$

Dies besagt, daß ein binärer Baum entweder ein Blatt (externer Knoten) ist, oder eine Wurzel zusammen mit einem linken und einem rechten Teilbaum, welche selbst binäre Bäume sind. Diese formale Gleichung kann übersetzt werden in eine Gleichung für die erzeugende Funktion $B(z)$:

$$B(z) = 1 + z B^2(z);$$

$$B(z) = \frac{1 - \sqrt{1 - 4z}}{2z} = \sum_{n \geq 0} \frac{1}{n+1} \binom{2n}{n}.$$

Binäre Bäume können verwendet werden, um arithmetische Ausdrücke darzustellen:



Die Auswertung kann mit Hilfe von Registern erfolgen; diese werden verwendet, um Zwischenergebnisse zu speichern.

Die Registerfunktion $Reg(t)$ ist definiert als die minimale Anzahl von Registern zur Auswertung des Baumes t unter Verwendung der optimalen Strategie. Es gilt

$$Reg(\square) = 0$$

$$Reg\left(\begin{array}{c} \circ \\ / \quad \backslash \\ t_1 \quad t_2 \end{array}\right) = \begin{cases} 1 + Reg(t_1) & \text{falls } Reg(t_1) = Reg(t_2) \\ \max\{Reg(t_1), Reg(t_2)\} & \text{sonst} \end{cases}$$

Es sei $\mathcal{R}_p(\mathcal{L}_p)$ die Familie der Bäume mit Registerfunktion $=p$ ($\geq p$);

$R_p(z)$ und $S_p(z)$ seien die entsprechenden erzeugenden Funktionen.

Man sieht unmittelbar:

$$\begin{aligned}
 R_p &= \begin{array}{c} \circ \\ \swarrow \quad \searrow \\ R_{p-1} \quad R_{p-1} \end{array} + \sum_{j < p} \begin{array}{c} \circ \\ \swarrow \quad \searrow \\ R_j \quad R_p \end{array} + \sum_{j < p} \begin{array}{c} \circ \\ \swarrow \quad \searrow \\ R_p \quad R_j \end{array} \\
 S_p &= \begin{array}{c} \circ \\ \swarrow \quad \searrow \\ S_{p-1} \quad S_{p-1} \end{array} + \begin{array}{c} \circ \\ \swarrow \quad \searrow \\ S_p \quad B \setminus S_{p-1} \end{array} + \begin{array}{c} \circ \\ \swarrow \quad \searrow \\ B \setminus S_{p-1} \quad S_p \end{array}
 \end{aligned}$$

Daher gilt

$$R_p = zR_{p-1}^2 + 2zR_p \sum_{j < p} R_j, \quad R_0 = 1$$

$$S_p = zS_{p-1}^2 + 2zS_p(B - S_{p-1}), \quad S_0 = B;$$

$$S_p = \frac{zS_{p-1}^2}{1 - 2zB + 2zS_{p-1}}.$$

Wir setzen $\epsilon := \sqrt{1-4z} = 1-2zB$:

$$\frac{1}{S_p} = \frac{\epsilon}{z} \frac{1}{S_{p-1}^2} + 2 \frac{1}{S_{p-1}}.$$

Wir multiplizieren mit ϵ/z und substituieren $U_p := \epsilon/zS_p$:

$$U_p = U_{p-1}^2 + 2U_{p-1} \Rightarrow U_p + 1 = (U_{p-1} + 1)^2 \Rightarrow$$

$$U_p = -1 + (U_0 + 1)^{2^p}; \quad U_0 = \frac{\epsilon}{zB} \Rightarrow$$

$$S_p = \frac{\epsilon}{z} \cdot \frac{1}{-1 + \left(\frac{1+\epsilon}{1-\epsilon}\right)^{2^p}}.$$

Es ist günstig, $z = u/(1+u)^2$ zu setzen:

$$S_p = \frac{1-u^2}{u} \cdot \frac{u^{2^p}}{1-u^{2^p}};$$

$$R_p = S_p - S_{p+1} = \frac{1-u^2}{u} \cdot \frac{u^{2^p}}{1-u^{2^{p+1}}}.$$

Sei nun M_n die mittlere Registerfunktion eines binären Baumes mit

n Knoten. Es gilt:

$$M_n = \frac{\sum_{p \geq 0} p \cdot [z^n] R_p(z)}{[z^n] B(z)} = \frac{[z^n] \sum_{p \geq 1} S_p(z)}{[z^n] B(z)} .$$

Wir studieren den Zähler:

$$\begin{aligned} E(z) &= \sum_{p \geq 1} S_p(z) = \frac{1-u^2}{u} \sum_{p \geq 1} \frac{u^{2p}}{1-u^{2p}} \\ &= \frac{1-u^2}{u} \sum_{\substack{p \geq 1 \\ \lambda \geq 1}} u^{2p\lambda} \\ &= \frac{1-u^2}{u} \sum_{k \geq 1} v_2(k) u^k \end{aligned}$$

mit $v_2(k) = \sum_{\substack{k=2^p \lambda, \\ p, \lambda \geq 1}} 1 \cdot (v_2(2n) = 1+v_2(n), v_2(2n+1)=0).$

Um den Koeffizienten von z^n in $E(z)$ zu finden, bedienen wir uns einer analytischen Methode. Das Verhalten der Koeffizienten wird nämlich weitgehend beschrieben durch das Verhalten der erzeugenden Funktion in der Nähe der Singularitäten am Konvergenzkreis. Es liegt eine Singularität bei $u=1$ vor.

Mit Hilfe der Mellintransformation findet man für $t \rightarrow 0$:

$$\sum_{k \geq 1} v_2(k) e^{-tk} \sim \sum_s \text{Res} \left(\frac{\Gamma(s) \zeta(s)}{2^s - 1} t^{-s} \right)$$

und insgesamt ($z \rightarrow 1/4$)

$$\begin{aligned} E(z) &\sim 2\epsilon \log_2 \epsilon + 2 \left(\frac{3}{2} - \log_2 2\pi + \frac{\gamma}{\log 2} \right) \\ &\quad + 4 \sum_{k \neq 0} \epsilon^{1-x_k} \quad \text{mit } x_k = \frac{2k\pi i}{\log 2} . \end{aligned}$$

Beachtet man noch, daß

$$[z^n] B(z) \sim \frac{4^n}{\sqrt{\pi}} n^{-3/2}$$

ist, folgt nun

$$\begin{aligned} M_n &\sim \log_4 n - \frac{1}{2} - \frac{\gamma}{2 \log 2} - \frac{1}{\log 2} + \log_2 2\pi \\ &\quad + \frac{1}{\log 2} \sum_{k \neq 0} \zeta(x_k) \Gamma\left(\frac{x_k}{2}\right) (x_k - 1)^n x_k^{-2} . \end{aligned}$$

Es gilt $n^{x_k/2} = e^{2k\pi i \cdot \log_4 n}$, und somit

$$M_n \sim \log_4 n + D(\log_4 n).$$

Hier ist $D(x)$ eine periodische Funktion mit Periode 1.

Die Fourierreihenentwicklung $D(x) = \sum_k d_k e^{2k\pi i x}$ ist:

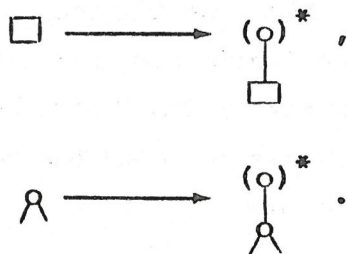
$$d_0 = -\frac{1}{2} - \frac{\gamma}{2 \log 2} - \frac{1}{\log 2} + \log_2 2\pi,$$

$$d_k = \frac{1}{\log 2} \zeta(x_k) \Gamma\left(\frac{x_k}{2}\right) (x_k - 1), \quad k \neq 0, \quad x_k = \frac{2k\pi i}{\log 2}.$$

Man kann auch *Motzkinbäume* betrachten:

$$\mathcal{M} = \square + \begin{array}{c} \circ \\ | \\ \mathcal{M} \end{array} + \begin{array}{c} \circ \\ / \quad \backslash \\ \mathcal{M} \quad \mathcal{M} \end{array}$$

\mathcal{M} kann aus \mathcal{B} durch Substitution gewonnen werden:



Daher:

$$M(z) = \frac{1}{1-z} B\left(\frac{z}{(1-z)^2}\right).$$

Die Registerfunktion von Bäumen aus \mathcal{M} wird durch die folgende Erweiterung gewonnen:

$$\text{Reg}\left(\begin{array}{c} \circ \\ | \\ t \end{array}\right) = \text{Reg}(t).$$

Es gilt auch

$$R_P^{(\mathcal{M})}(z) = \frac{1}{1-z} R_P\left(\frac{z}{(1-z)^2}\right) \text{ usw.}$$

Deshalb ist die mittlere Registerfunktion jetzt gegeben durch

$$\frac{[z^n] \frac{1}{1-z} E\left(\frac{z}{(1-z)^2}\right)}{[z^n] \frac{1}{1-z} B\left(\frac{z}{(1-z)^2}\right)}$$

Ähnlich wie vorher erhält man

$$\sim \log_4 n + \bar{D}(\log_4 n).$$

Die Situation kann noch verallgemeinert werden durch Gewichte:

$$\mathcal{F} = c_0 \square + c_1 \cdot \begin{array}{c} \circ \\ | \\ \mathcal{F} \end{array} + c_2 \begin{array}{c} \circ \\ / \quad \backslash \\ \mathcal{F} \quad \mathcal{F} \end{array} .$$

* ————— *

In einem binären Baum bezeichnen wir einen Knoten t als *kritisch*, falls die Registerfunktion des linken und rechten Unterbaumes gleich sind. Nach längerer Rechnung erhält man: die mittlere Anzahl der kritischen Knoten eines binären Baumes mit n Knoten ist

$$\frac{[z^n]V(z)}{[z^n]B(z)},$$

wo

$$V(z) = \frac{u(1+u)}{1-u} - 2 \cdot \frac{1-u^2}{u} \sum_{k \geq 1} \omega(k) u^k ;$$

$$\omega(k) = i \text{ falls } k = 2^m(1+2i).$$

Durch ähnliche Methoden wie vorher erhält man:

"ungefähr $\frac{n}{3}$ Knoten sind kritisch".

Man kann auch explizit errechnen, daß

$$[z^n]V(z) = \binom{2n}{n-1} - 2 \sum_{k \geq 1} \omega(k) \left[\binom{2n}{n+1-k} - 2 \binom{2n}{n-k} + \binom{2n}{n-1-k} \right].$$

Durch partielle Summation (zweimal durchzuführen) und Information über

$$\sum_{n < N} \sum_{k \leq n} \omega(k)$$

kann man das asymptotische Verhalten auch relativ "elementar" (dafür mühsam) errechnen.

Auf eine ungelöste Frage im Zusammenhang mit der Registerfunktion

binärer Bäume soll noch eingegangen werden:

Die erzeugende Funktion der binären Bäume mit Registerfunktion $\langle p \rangle$ ist

$$B(z) - S_p(z) = \frac{F_{2^{p-1}}(z)}{F_{2^p}(z)},$$

mit

$$F_n(z) = \sum_j \binom{n-1-j}{j} (-z)^j \quad (\text{Fibonaccipolynom}).$$

Die linksseitige Höhe h eines binären Baumes ist definiert durch

$$h(\square) = 0$$

$$h\left(\begin{array}{c} \circ \\ / \quad \backslash \\ t_1 \quad t_2 \end{array}\right) = \max\{1+h(t_1), h(t_2)\}.$$

Die erzeugende Funktion der Bäume mit $h(t) < h$ ist

$$\frac{F_h(z)}{F_{h+1}(z)}.$$

Somit ist die Anzahl der binären Bäume mit n Knoten und Registerfunktion $\langle p \rangle$ gleich der Anzahl der binären Bäume mit n Knoten und linksseitiger Höhe $< 2^{p-1}$. Gesucht ist eine geeignete Bijektion dieser Objekte.

LITERATUR.

- [1] N.G.de Bruijn, D.E.Knuth, S.O.Rice: The average height of planted plane tree, in: Graph Theory and Computing (R.C.Read,Ed.), 15-22, Academic Press, 1972.
- [2] P.Flajolet: Analyse d'algorithmes de manipulation d'arbres et de fichiers, Thèse Paris-Sud-Orsay, 1979.
- [3] P.Flajolet, A.Odlyzko: The average height of binary trees and other simple trees, JCSS 25(1982), 171-213.
- [4] P.Flajolet, H.Prodinger: The number of registers to evaluate unary-binary trees, preprint.
- [5] P.Flajolet, J.C.Raoult, J.Vuillemin: The number of registers

required for evaluating arithmetical expressions,
TCS 9(1979), 99-125.

- [6] R.Kemp: The average number of registers to evaluate a binary tree optimally, Acta Inf. 11(1979), 313-372.
- [7] H.Prodinger: Some results relative to the register function of a binary tree, preprint.
- [8] H.Prodinger, R.F.Tichy: Über zahlentheoretische Funktionen, die mit der Ziffernsumme zusammenhängen, preprint.