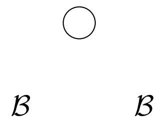
ABZÄHLPROBLEME BEI BÄUMEN

HELMUT PRODINGER (WIEN)

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

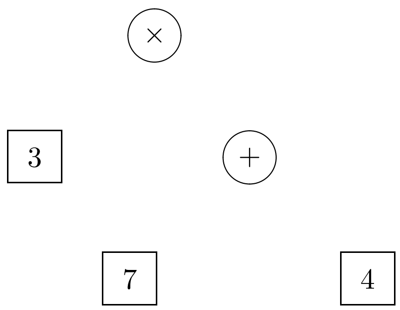
$$\mathcal{B} \quad = \quad \square \qquad +$$



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 $B(z)$:

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

Binäre Bäume können verwendet werden, um arithmetische Ausdrücke darzustellen:



$$\longleftrightarrow \qquad 3 \times (7 + 4)$$

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:

$$\mathrm{Reg}(\Box) = 0$$

$$\mathrm{Reg}\left(\begin{array}{ccc} & \bigcirc & \\ t_1 & & t_2 \end{array}\right) = \begin{cases} 1 + \mathrm{Reg}(t_1) & \text{falls } \mathrm{Reg}(t_1) = \mathrm{Reg}(t_2) \\ \max\{\mathrm{Reg}(t_1), \mathrm{Reg}(t_2)\} & \text{sonst} \end{cases}$$

Es sei $\mathcal{R}_p$ $(\mathcal{S}_p)$ die Familie der Bäume mit Registerfunktion $= p$ $(\geq p)$; $R_p(z)$ und $S_p(z)$ seien die entsprechenden erzeugenden Funktionen. Man sieht unmittelbar:

$$\mathcal{R}_p = \begin{array}{ccc} & \bigcirc & \\ \mathcal{R}_{p-1} & & \mathcal{R}_{p-1} \end{array} + \begin{array}{ccc} & \bigcirc & \\ \sum\limits_{j<p} \mathcal{R}_j & & \mathcal{R}_p \end{array} + \begin{array}{ccc} & \bigcirc & \\ \mathcal{R}_p & & \sum\limits_{j<p} \mathcal{R}_j \end{array}$$

$$\mathcal{S}_p = \begin{array}{ccc} & \bigcirc & \\ \mathcal{S}_{p-1} & & \mathcal{S}_{p-1} \end{array} + \begin{array}{ccc} & \bigcirc & \\ \mathcal{B} \setminus \mathcal{S}_{p-1} & & \mathcal{S}_p \end{array} + \begin{array}{ccc} & \bigcirc & \\ \mathcal{S}_p & & \mathcal{B} \setminus \mathcal{S}_{p-1} \end{array}$$

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 $\varepsilon := \sqrt{1-4z} = 1 - 2zB$:

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

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

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

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

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

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:

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

Wir studieren den Zähler:

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

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

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 *Mellin–Transformation* findet man für $t \to 0$:

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

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

$$E(z) \sim 2\varepsilon \log_2 \varepsilon + 2\left(\frac{3}{2} - \log_2 2\pi + \frac{\gamma}{\log 2}\right) + 4 \sum_{k\ne 0} \varepsilon^{1-\chi_k} \quad \text{mit } \chi_k = \frac{2k\pi i}{\log 2}$$

Beachtet man noch, daß

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

ist, folgt nun

$$M_n \sim \log_4 n - \frac{1}{2} - \frac{\gamma}{2\log 2} - \frac{1}{\log 2} + \log_2 2\pi + \frac{1}{\log 2} \sum_{k\ne 0} \zeta(\chi_k)\Gamma(\frac{\chi_k}{2})(\chi_k - 1) n^{\chi_k/2} .$$

Es gilt $n^{\chi_k/2} = e^{2k\pi i \cdot \log_4 n}$, 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:

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

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

Man kann auch *Motzkinbäume* betrachten:

$$\mathcal{M} \quad = \quad \square \; + \; \begin{array}{c}\bigcirc \\ \\ \mathcal{M}\end{array} \; + \; \begin{array}{c}\bigcirc \\ \\ \mathcal{M}\;\;\mathcal{M}\end{array}$$

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

$$\square \quad \mapsto \quad \begin{array}{c}(\;\bigcirc\;)^* \\ \\ \square\end{array}$$

$$\bigcirc \quad \mapsto \quad \begin{array}{c}(\;\bigcirc\;)^* \\ \\ \bigcirc\end{array}$$

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:

$$\operatorname{Reg}\left(\begin{array}{c}\bigcirc \\ \\ t\end{array}\right) = \operatorname{Reg}(t)$$

Es gilt auch

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

Deshalb ist die mittlere Registerfunktion jetzt gegeben durch

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

Ähnlich wie vorher erhält man

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

$$\mathcal{F} \;=\; c_0 \cdot \; \square \quad + c_1 \cdot \begin{array}{c} \bigcirc \\ \\ \mathcal{F} \end{array} \quad + c_2 \cdot \begin{array}{ccc} & \bigcirc & \\ & & \\ \mathcal{F} & & \mathcal{F} \end{array}$$

Die Situation kann noch verallgemeinert werden durch Gewichte:

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 $< p$ ist

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

mit

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

Die *linksseitige Höhe* h eines binären Baumes ist definiert durch

$$h(\square) = 0$$

$$h\left(\begin{array}{ccc} & \bigcirc & \\ & & \\ t_1 & & t_2 \end{array} \right) \quad = \quad \max\{1 + h(t_1), h(t_2)\}$$

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 $< p$ 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 trees, 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, SIAM J. Computing 15(1986), 629-640.
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 recent results on the register function of a binary tree, Annals of Discrete Math. 33(1987), 241-260.
8. H.Prodinger, R.F.Tichy: Über ein zahlentheoretisches Problem aus der Informatik, Sitzungsberichte der Österreichischen Akademie der Wissenschaften, Abt. II, 192(1983), 385-396.