Binäre Suchbäume

Wolfgang Mulzer

1 Einfache Binäre Suchbäume

1.1 Darstellung

Ein binärer Suchbaum besteht aus Knoten. Ein Knoten `n` hat folgende Felder:

- `k`, `v`: Schlüssel und Wert
- `parent`: Elternknoten
- `left`, `right`: linkes und rechtes Kind

1.2 Nachschlagen

```
get(k)
  n <- root
  while n != NULL do
    if n.k == k then
      return n.v
    if k < n.k then
      n <- n.left
    else
      n <- n.right
  throw NoSuchElementException
```

1.3 Einfügen

```
put(k,v)
  if root == NULL then
    root <- new node for (k,v)
    return
  n <- root
  while true do
    if n.k == k then
      n.v <- v
      return
    if k < n.k then
      if n.left != NULL then
        n <- n.left
```

```
      else
        n.left <- new node for (k,v)
        return
    else
      if n.right != NULL then
        n <- n.right
      else
        n.right <- new node for (k,v)
        return
```

1.4 Löschen

```
remove(k)
  // lokalisiere den Knoten fuer k
  n <- root
  while n != NULL && n.k != k do
    if k < n.k then
      n <- n.left
    else
      n <- n.right
  if n == NULL then
    throw NoSuchElementException
  // wenn der Knoten n kein linkes oder kein rechtes Kind hat, loesche ihn direkt
  if n.left == NULL then
    // ersetze n durch das rechte Kind. Wir zeigen die noetigen Zeigeroperationen
    // nur einmal
    if n.right != NULL then
      n.right.parent <- n.parent
    if n.parent != NULL then
      if n.parent.left == n then
        n.parent.left <- n.right
      else
        n.parent.right <- n.right
    return
  else if n.right == NULL then
    replace n by its left child
    return
  // n hat beide Kinder -> finde den Vorgaenger von n
  m <- n.left
  while m.right != NULL do
    m <- m.right
  // nun ist m der Vorgaenger von n, m hat kein rechtes Kind
  replace m by its left child
  replace n by m
```

2 AVL-Bäume

Ein *AVL-Baum* ist ein *höhen-balancierter* binärer Suchbaum. Das bedeutet, dass sich an jedem Knoten v des AVL-Baums die Höhen des linken und des rechten Teilbaums höchstens um 1 unterscheiden. Wir wollen nun sehen, dass diese Eigenschaft einen fast perfekten Baum garantiert.

2.1 Die Höhe von AVL-Bäumen

Theorem 2.1. *Ein AVL-Baum mit n Knoten hat Höhe $O(\log n)$.*

Proof. Für $h \in \{0, 1, \dots, \}$ sei F_h definiert als die *minimale* Anzahl von Knoten, die ein AVL-Baum der Höhe h besitzen kann. Dann gilt

$$F_h > F_{h-1} \qquad \text{für alle } h \in \{1, \dots\}, \tag{1}$$

da ein Baum der Höhe h immer einen echten Teilbaum der Höhe $h-1$ enthält. Außerdem ist

$$F_0 = 1, \tag{2}$$

denn es gibt nur einen Baum der Höhe 0: der Baum der nur aus der Wurzel besteht. Außerdem haben wir

$$F_1 = 2, \tag{3}$$

den von den drei möglichen binären Baumen der Höhe 1 sind die beiden Bäume mit zwei Knoten AVL-Bäume. Schließlich gilt für $h \geq 2$:

$$F_h = F_{h-2} + F_{h-1} + 1, \tag{4}$$

denn ein AVL-Baum der Höhe h besteht aus einer Wurzel und zwei Teilbäumen, deren Höhen sich um höchstens 1 unterscheiden. Da die Teilbäume unabhängig sind, muss die Anzahl der Knoten in den Teilbäumen minimal sein, damit die Anzahl der Knoten im gesamten Baum minimal ist. Wegen (1) muss einer der Teilbäume Höhe $h-2$ haben.

Wir beweisen durch Induktion: Für alle $h \in \{0, 1, \dots, \}$ gilt

$$F_h \geq \left(\frac{3}{2}\right)^h. \tag{5}$$

Die Aussage ist klar für $h \in \{0, 1\}$, denn nach (2, 3) haben wir

$$F_0 = 1 \geq \left(\frac{3}{2}\right)^0 \text{ und } F_1 = 2 \geq \left(\frac{3}{2}\right)^1.$$

Sei nun $h \geq 2$. Nach Induktionsannahme und (4) gilt:

$$\begin{aligned}
F_h = F_{h-2} + F_{h-1} + 1 &\geq \left(\frac{3}{2}\right)^{h-2} + \left(\frac{3}{2}\right)^{h-1} + 1 \\
&= \left(\frac{3}{2}\right)^{h-2} \left(1 + \frac{3}{2}\right) + 1 \\
&> \left(\frac{3}{2}\right)^{h-2} \cdot \frac{9}{4} \\
&= \left(\frac{3}{2}\right)^{h}.
\end{aligned}$$

Damit ist (5) bewiesen.

Betrachte nun einen AVL-Baum T mit n Knoten, und nimm an, dass T die Höhe h besitzt. Dann gilt nach Definition von F_h, dass $n \geq F_h$ ist. Nach (5) ist außerdem $F_h \geq (3/2)^h$. Also gilt $n \geq (3/2)^h$, oder äquivalent $h \leq \log_{3/2} n \approx 1.71 \log_2 n = O(\log n)$, wie behauptet. □

Anmerkung. Durch exaktes Lösen der Rekursionsgleichung (2, 3, 4) erhält man sogar, dass $h \leq \log_\varphi(n+2) \approx 1.44 \log(n+2)$ ist, wobei $\varphi = (1+\sqrt{5})/2$ den goldenen Schnitt bezeichnet.

Das wollen wir nun beweisen. Dazu verwenden wir die Technik der *erzeugenden Funktionen*. Wir stellen die Folge F_h als Funktion dar, und benutzen algebraische Manipulationen, um eine explizite Darstellung der Folgenglieder zu erhalten. Wir schreiben also

$$F(x) = \sum_{h=0}^{\infty} F_h x^h.$$

Dann gilt:

$$\begin{aligned}
F(x) &= 1 + 2x + \sum_{h=2}^{\infty} F_h x^h && \text{(nach (2, 3))}\\
&= 1 + 2x + \sum_{h=2}^{\infty} (F_{h-2} + F_{h-1} + 1) x^h && \text{(nach (4))}\\
&= 1 + 2x + \sum_{h=2}^{\infty} F_{h-2} x^h + \sum_{h=2}^{\infty} F_{h-1} x^h + \sum_{h=2}^{\infty} x^h && \text{(Aufteilen der Summanden)}\\
&= 1 + 2x + x^2 \left(\sum_{h=0}^{\infty} F_h x^h \right) + x \left(\sum_{h=1}^{\infty} F_h x^h \right) + \left(\sum_{h=0}^{\infty} x^h \right) - 1 - x && \text{(Indexverschiebung)}\\
&= 1 + 2x + x^2 F(x) + x\,(F(x) - 1) + \left(\sum_{h=0}^{\infty} x^h \right) - 1 - x && \text{(Definition von } F(x)\text{)}\\
&= x^2 F(x) + xF(x) + \frac{1}{1-x}. && \text{(Vereinfachen, geometrische Reihe)}
\end{aligned}$$

Auflösen nach $F(x)$ ergibt die folgende Funktionalgleichung:

$$F(x) = \frac{1}{(x^2 + x - 1)(x-1)}.$$

Schreibe nun $\varphi = (1+\sqrt{5})/2$ und $\widehat{\varphi} = (1-\sqrt{5})/2$. Wie man schnell nachrechnet, haben die beiden Zahlen φ und $\widehat{\varphi}$ die folgenden Eigenschaften:

$$\text{(i) } \varphi + \widehat{\varphi} = 1; \quad \text{(ii) } \varphi - \widehat{\varphi} = \sqrt{5}; \quad \text{(iii) } \varphi \cdot \widehat{\varphi} = -1; \quad \text{(iv) } \varphi^2 = \varphi + 1; \text{ und (v) } \widehat{\varphi}^2 = \widehat{\varphi} + 1.$$

Aus (i, iii) folgt, dass $(x + \widehat{\varphi})(x + \varphi) = x^2 + x - 1$ ist. Also:

$$F(x) = \frac{1}{(x + \widehat{\varphi})(x + \varphi)(x-1)}.$$

Das Ziel ist nun, aus der Funktionalgleichung für $F(x)$ eine explizite Darstellung der Folgenglieder F_h zu gewinnen. Dazu wollen wir die Funktionalgleichung auf die geometrische Reihe $1/(1-ax) = \sum_{h=0}^{\infty} a^h x^h$ zurückführen. Das geeignete Werkzeug zu diesem Zweck ist die bekannte *Partialbruchzerlegung*. Wir machen den Ansatz

$$F(x) = \frac{1}{(x + \widehat{\varphi})(x + \varphi)(x-1)} = \frac{\alpha}{x + \widehat{\varphi}} + \frac{\beta}{x + \varphi} + \frac{\gamma}{x-1},$$

wobei α, β, γ gesucht sind. Wenn wir diese Gleichung mit $x + \widehat{\varphi}$ multiplizieren und $x = -\widehat{\varphi}$ setzen, so erhalten wir wegen (ii, v, iii)

$$\alpha = \frac{1}{(-\widehat{\varphi} + \varphi)(-\widehat{\varphi} - 1)} = \frac{1}{\sqrt{5}(-\widehat{\varphi}^2)} = -\frac{\varphi^2}{\sqrt{5}}.$$

Analog sehen wir mit (ii, iv, iii), dass

$$\beta = \frac{1}{(-\varphi + \widehat{\varphi})(-\varphi - 1)} = \frac{1}{-\sqrt{5}(-\varphi^2)} = \frac{\widehat{\varphi}^2}{\sqrt{5}}$$

und mit (iv, v, iii), dass

$$\gamma = \frac{1}{(1+\varphi)(1+\widehat{\varphi})} = \frac{1}{(\varphi\widehat{\varphi})^2} = 1$$

ist. Nun können wir mit der Partialbruchzerlegung für $F(x)$ wie geplant arbeiten.

$$\begin{aligned}
F(x) &= -\frac{\varphi^2}{\sqrt{5}} \cdot \frac{1}{x+\widehat{\varphi}} + \frac{\widehat{\varphi}^2}{\sqrt{5}} \cdot \frac{1}{x+\varphi} + \frac{1}{x-1} \\
&= -\frac{\varphi^2}{\sqrt{5}} \cdot \frac{1}{\widehat{\varphi}(1+x/\widehat{\varphi})} + \frac{\widehat{\varphi}^2}{\sqrt{5}} \cdot \frac{1}{\varphi(1+x/\varphi)} - \frac{1}{1-x} && \text{(Ausklammern)} \\
&= \frac{\varphi^2}{\sqrt{5}} \cdot \frac{\varphi}{1-\varphi x} - \frac{\widehat{\varphi}^2}{\sqrt{5}} \cdot \frac{\widehat{\varphi}}{1-\widehat{\varphi}x} - \frac{1}{1-x} && \text{(wegen (iii))} \\
&= \frac{\varphi^3}{\sqrt{5}} \cdot \sum_{h=0}^{\infty} \varphi^h x^h - \frac{\widehat{\varphi}^3}{\sqrt{5}} \cdot \sum_{h=0}^{\infty} \widehat{\varphi}^h x^h - \sum_{h=0}^{\infty} x^h - \frac{1}{1-x} && \text{(geometrische Reihe)} \\
&= \sum_{h=0}^{\infty} \left(\frac{\varphi^{h+3} - \widehat{\varphi}^{h+3}}{\sqrt{5}} - 1 \right) x^h. && \text{(Zusammenfassen)}
\end{aligned}$$

Durch Koeffizientenvergleich können wir nun die Formel

$$F_h = \frac{1}{\sqrt{5}} \left(\varphi^{h+3} - \widehat{\varphi}^{h+3}\right) - 1,$$

für alle $h \geq 0$, ableiten. Es gilt also

$$n \geq F_h \geq \frac{1}{\sqrt{5}} \varphi^{h+3} - 2$$

und somit

$$h \leq \log_\varphi(n+2) + \log_\varphi(\sqrt{5}) - 3,$$

wie behauptet.