Lehrstuhl für Informatik I
— Informationsstrukturen und
wissensbasierte Systeme —
Universität Würzburg

Prof. Dr. H. Noltemeier

Würzburg, 27. Mai 1998

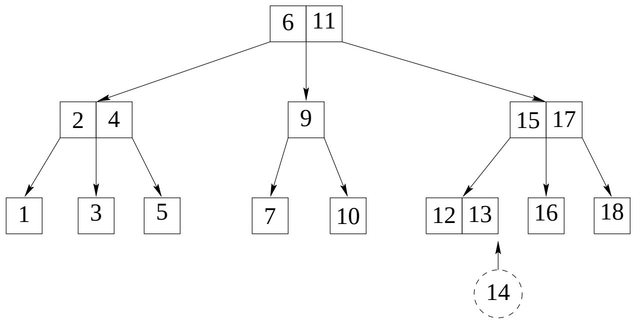
Übungen zur Vorlesung
Praktische Informatik II (SS98)

Musterlösung zum 1. Übungsblatt

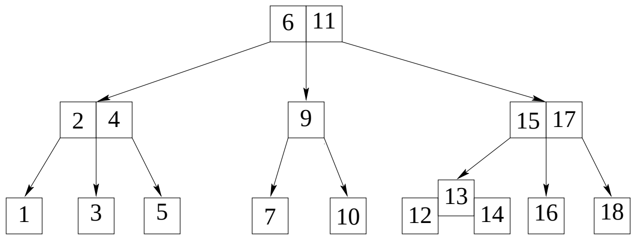
Lösung zu Aufgabe 1: (Operationen im B-Baum)

In dem gegebenen B-Baum vom Typ $(1, \cdot)$ war ein Folge von Einfügen- und Entfernen-Operationen durchzuführen.

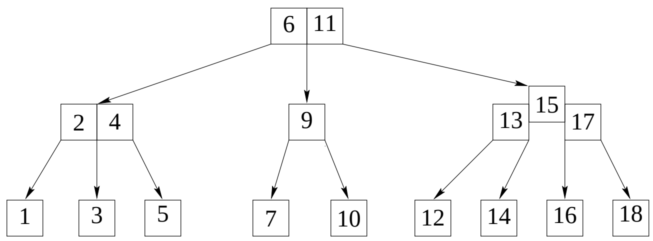
(a) Einfügen von 14.



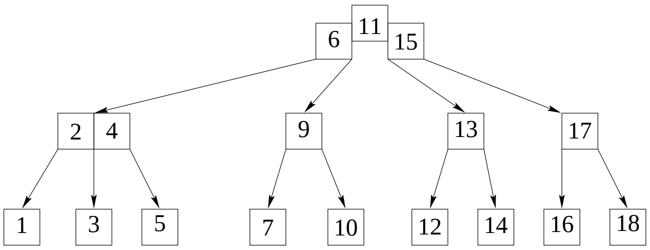
Die Suche nach dem Schlüsselwert 14 endet erfolglos in einem Blatt. An dieser Stelle wird der neue Schlüsselwert eingefügt.



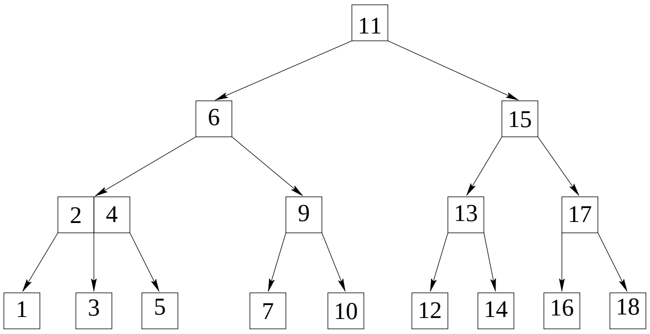
Da das Blatt bereits vor dem Einfügen des neuen Schlüsselwertes voll war, muß es gesplittet werden. Der mittlere Schlüsselwert (13) wird in den Vaterknoten eingefügt.



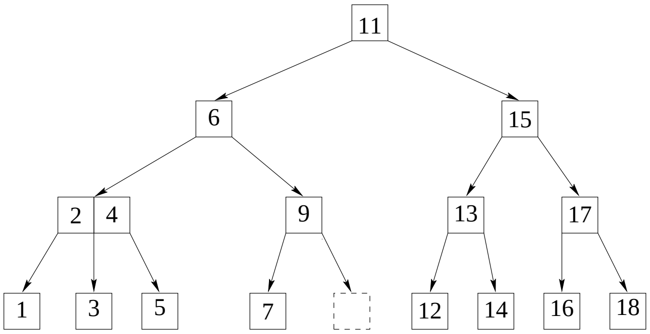
Nun läuft dieser Knoten ebenfalls über. Er wird gesplittet, der Schlüsselwert 15 wird in die Wurzel eingefügt.



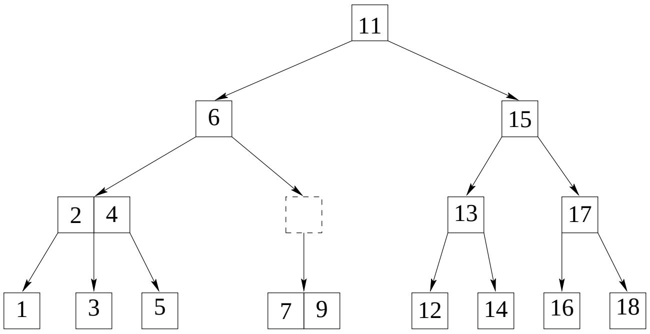
Die Wurzel läuft über und wird gesplittet. Der mittlere Schlüsselwert gelangt in einen neu erzeugten Wurzelknoten, die Höhe des Baumes wächst um 1.



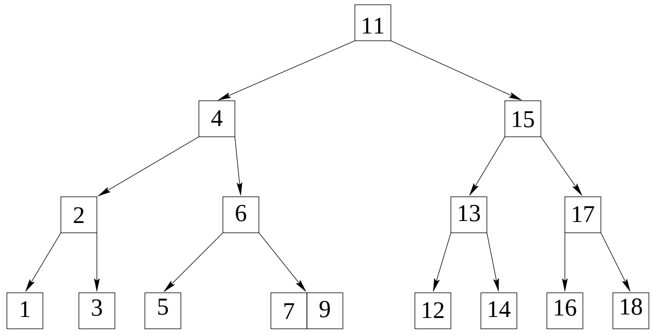
(b) Löschen von 10.



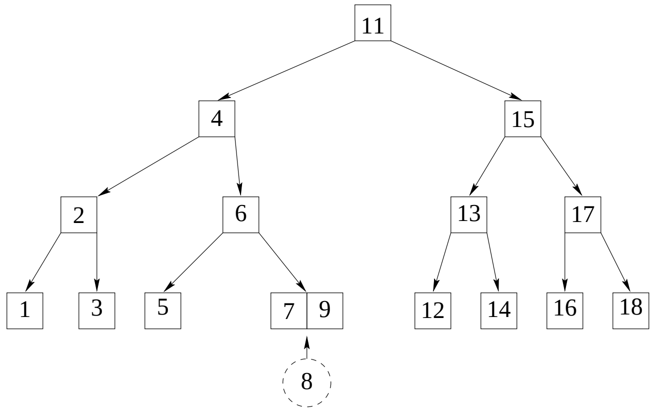
Der Schlüsselwert 10 steht in einem Blatt und kann direkt entfernt werden, was einen Unterlauf in diesem Knoten erzeugt. Da keine Geschwisterknoten mit mehr als $k = 1$ Schlüsselwerten existieren, wird einer der Geschwisterknoten mit dem Trennschlüssel 9 aus dem Vaterknoten verkettet.



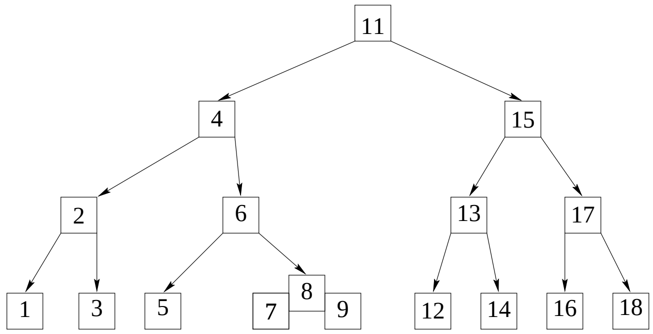
Durch diese Verkettung ergibt sich nun ein Unterlauf im Vaterknoten. Dieser hat jedoch einen Bruderknoten mit 2 Schlüsselwerten $\{2, 4\}$, also wird die Operation durch eine Schlüsselverschiebung über den gemeinsamen Vaterknoten abgeschlossen.



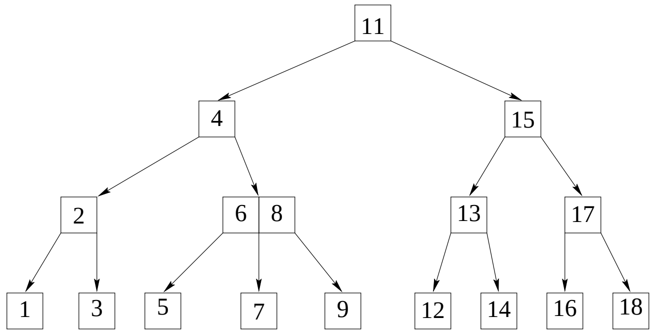
(c) Einfügen von 8.



Die Suche nach dem Schlüsselwert 8 endet erfolglos in einem Blatt. An dieser Stelle wird der neue Schlüsselwert eingefügt.



Das das Blatt vor dem Einfügen des neuen Schlüsselwertes bereits voll war, muß es gesplittet werden. Der mittlere Schlüsselwert (8) wird in den Vaterknoten eingefügt. Dieser enthält nun genau $2k$ Schlüsselwerte, also ist die Operation abgeschlossen.



Lösung zu Aufgabe 2: (Höhe von B-Bäumen)

In einen B-Baum vom Typ $(k, \cdot)$, welcher anfangs die Menge $\{1\}$ repräsentiert, werden der Reihe nach die Schlüsselwerte $2, 3, 4, \dots$ eingefügt. Bei welchem Schlüsselwert wird zum ersten Mal die Höhe $h \in \mathbb{N}$ erreicht?

Wir definieren eine Funktion $n(h)$, die für eine vorgegebene Höhe h den Schlüsselwert $n(h)$ liefert, bei dem zum ersten Mal die Höhe h erreicht wird:

$$n(h) := 1 + 2((k+1)^h - 1)$$

Diese Behauptung beweisen wir durch vollständige Induktion nach h:

Induktionsanfang: $h = 1$

Es gilt: $n(1) = 1 + 2((k+1)^1 - 1) = 1 + 2k$. Da die Wurzel höchstens $2k$ Schlüsselwerte aufnehmen kann, läuft sie beim $(1+2k)$-ten über, und die Aussage ist für $h = 1$ gezeigt.

Induktionsvoraussetzung: Die Aussage gelte für ein $h \in \mathbb{N}$.

Induktionsschritt: $h \to h+1$

Definiere $\Delta := n(h+1) - n(h)$, d.h. Δ ist genau die Anzahl der Schlüssel, die man in den Baum, der gerade eben die Höhe h erreicht hat, einfügen muß, damit er zur Höhe $h+1$ überläuft.

Da die Schlüsselwerte streng monoton wachsen, wird stets im rechtesten Blatt eingefügt, so auch der Schlüsselwert $n(h)$. Mit $n(h)$ ist aber die Wurzel übergelaufen, d.h. es haben Aufsplittungen stattgefunden, die sich bis in die Wurzel fortgesetzt haben. Also existiert ein Pfad von der Wurzel zum rechtesten Blatt, auf dem jeder Knoten außer der Wurzel k Schlüssel enthält, die (neu erzeugte) Wurzel enthält einen Schüssel. Genauer: Der Pfad enthält genau h Knoten mit jeweils k Schlüsseln und einen Knoten mit einem Schlüssel.

Da neue Schlüssel stets im rechtesten Blatt eingefügt werden, müssen genau die Knoten auf dem Pfad von der Wurzel zum rechtesten Blatt zum Überlaufen gebracht werden. D.h. man braucht $k+1$ Schlüssel für das Blatt, $(k+2)^2$ für den Vater des Blattes etc., und schließlich $2k(k+1)^h$ Schlüsselwerte, um die Wurzel überlaufen zu lassen. Also gilt $\Delta = 2k(k+1)^h$, und mit der Induktionsvoraussetzung

$$\begin{aligned} n(h) &= 1 + 2((k+1)^h - 1) \\ &= 1 + 2k(1 + (k+1) + (k+1)^2 + \cdots + (k+1)^{h-1}) \end{aligned}$$

folgt schließlich:

$$\begin{aligned} n(h) + \Delta &= 1 + 2k(1 + (k+1) + (k+1)^2 + \cdots + (k+1)^{h-1}) + 2k(k+1)^h \\ &= 1 + 2k(1 + (k+1) + (k+1)^2 + \cdots + (k+1)^{h-1} + (k+1)^h) \\ &= 1 + 2((k+1)^{h+1} - 1) \\ &= n(h+1) \end{aligned}$$

Damit ist die Behauptung gezeigt. $\square$

Lösung zu Aufgabe 3: (Existenz von B-Bäumen)

(a) **Zu zeigen:**

Zu jeder Zahl $k \in \mathbb{N}$ und jeder endlichen nichtleeren Schlüsselwertmenge S gibt es einen B-Baum vom Typ $(k, \cdot)$, welcher die gegebene Schlüsselwertmenge repräsentiert.

Beweis:

Vollständige Induktion nach der Anzahl n der Schlüsselwerte in S.

(IA) Sei $n = 1$, $S = \{m\}$. Diese Schlüsselwertmenge wird von einem B-Baum vom Typ $(k, 0)$ repräsentiert, der nur aus der Wurzel besteht, die den einzigen Schlüsselwert m enthält.

(IV) Sei für n bereits gezeigt, daß es für jede Zahl $k \in \mathbb{N}$ und jede endliche Schlüsselwertmenge S mit n Elementen einen B-Baum vom Typ $(k, \cdot)$ gibt, welcher die gegebene Schlüsselwertmenge repräsentiert.

(IS) Sei S eine Schlüsselwertmenge mit $n + 1$ Elementen, d.h. man kann S darstellen als $S = S' \cup \{m'\}$ mit $|S'| = n$ und $|S| = n + 1$. Nach (IV) existiert ein B-Baum vom Typ $(k, \cdot)$ für jede endliche Schlüsselwertmenge mit n Elementen, d.h. insbesondere existiert auch für S' ein solcher Baum.

Durch Einfügen des Elementes m' in die Page, in der die Suche nach m' erfolglos abbrechen würde, und anschließendem rebalancieren der Pages auf dem Pfad zur Wurzel entsteht ein neuer Baum, der die Struktureigenschaften eines B-Baumes vom Typ $(k, \cdot)$ erfüllt.

Also kann die Schlüsselwertmenge S mit $n + 1$ Schlüsseln ebenfalls durch einen B-Baum vom Typ $(k, \cdot)$ repräsentiert werden, was den Beweis der Behauptung abschließt. $\square$

(b) **Zu zeigen:**

Zu jeder Zahl $k \in \mathbb{N}$ gibt es eine Zahl $n(k) \in \mathbb{N}$ mit der Eigenschaft: Zu jeder endlichen Schlüsselwertmenge S mit einer Mächtigkeit $|S| \geq n(k)$ gibt es einen B-Baum vom Typ $(k, \cdot)$, welcher die gegebene Schlüsselwertmenge repräsentiert und dessen Wurzel mindestens k Schlüsselwerte enthält.

Beweis:

Abschätzung der Anzahl der Pages. Sei $n_{h,max}$ die maximale Anzahl der Pages in einem Baum der Höhe h. Jede Page kann maximal $2k + 1$ Söhne haben, also gilt

$$n_{h,max} \leq 1 + (2k + 1) + (2k + 1)^2 + \cdots + (2k + 1)^h$$

Durch Umformen dieser geometrischen Reihe ergibt sich

$$n_{h,max} \leq \frac{1}{2k} \cdot ((2k + 1)^{h+1} - 1)$$

Analog kann man die minimale Anzahl $n_{h,min}$ der Pages eines Baumes der Höhe h abschätzen, unter der Bedingung, daß die Wurzel mindestens

k Schlüssel enthält:

$$\begin{aligned} n_{h,min} &\geq 1 + (k+1) + (k+2)^2 + \cdots + (k+1)^h \\ &= \frac{1}{k} \cdot ((k+1)^{h+1} - 1) \end{aligned}$$

Man berechnet die Höhe h eines Baumes, bei der die maximale Anzahl der aufzunehmenden Schlüsselwerte größer ist als die minimale Anzahl der aufznehmenden Schlüsselwerte bei einem Baum der Höhe $h+1$, mit der Einschränkung, daß die Wurzel mindestens k Schlüsselwerte enthält.

$$\begin{aligned} S_{h+1,min} &\leq S_{h,max} \\ k \cdot n_{h+1,min} &\leq 2k \cdot n_{h,max} \\ k \cdot \frac{1}{k}((k+1)^{h+2} - 1) &\leq 2k \cdot \frac{1}{2k}((2k+1)^{h+1} - 1) \\ (k+1)^{h+2} &\leq (2k+1)^{h+1} \qquad (1) \end{aligned}$$

Da $k \geq 1$, gilt $\frac{k}{2} \geq \frac{1}{2}$ und somit $2k+1 \geq \frac{3}{2}(k+1)$. Daraus ergibt sich, daß für obige Forderung (1) folgende Gleichung hinreichend ist:

$$\begin{aligned} (k+1)^{h+2} &\leq (\frac{3}{2}(k+1))^{h+1} \\ (k+1) &\leq (\frac{3}{2})^{h+1} \\ h+1 &\geq \log_{\frac{3}{2}+1}(k+1) \\ h &\geq \log_{\frac{3}{2}+1}(k+1) - 1 \end{aligned}$$

$$\begin{aligned} |S| &\geq k \cdot n \\ &\geq (k+1)^{h-1} - 1 \\ &\geq (k+1)^{\log_{\frac{3}{2}+1}(k+1)} - 1 \\ &=: n(k) \end{aligned}$$

Somit gibt es zu jeder endlichen Schlüsselwertmenge S mit einer Mächtigkeit

$$|S| \geq (k+1)^{\log_{\frac{5}{2}}(k+1)} - 1 =: n(k)$$

einen B-Baum vom Typ $(k, \cdot)$, welcher die gegebene Schlüsselwertmenge repräsentiert und dessen Wurzel mindestens k Schlüsselwerte enthält. □

Lösung zu Aufgabe 4: (Modifizierte B-Baum-Operationen)

Die Operationen für den B-Baum werden nun leicht modifiziert. Beim Einfügen wird ähnlich wie beim Löschen bei einem Überlauf nicht sofort der Knoten gesplittet, sondern es erfolgt zunächst ein Vergleich mit einem Bruder. Sind beide Knoten voll (sie enthalten jeweils $2k$ Schlüsselwerte), so wird die Vereinigung der beiden Schlüsselwertmengen auf drei Knoten verteilt, ein Schlüsselwert kommt in den Vaterknoten. Rekursiv wird mit einem evtl. Überlauf im Vaterknoten verfahren.

Durch diese Strategie kann erreicht werden, daß in jedem Knoten außer der Wurzel mindestens $\frac{4k}{3}$ Schlüsselwerte enthalten sind.

(a) Algorithmus 1 beschreibt eine DELETE-Operation, so daß die untere Schranke von $\frac{4k}{3}$ Schlüsselwerten gewahrt bleibt.

Algorithmus 1 DELETE x

Suche Schlüsselwert x im Baum
if Schlüsselwert x gefunden in Page p **then**
  Lösche Schlüsselwert x aus Page p
  if Page p ist kein Blatt **then**
    Finde kleinestes Element y des Teilbaumes unmittelbar rechts von x. $\{y$ steht in einem Blatt.$\}$
    Setze y an die Stelle von x
    $p \leftarrow$ Blatt, aus dem y entnommen wurde
  if Blatt p enthält weniger als $\frac{4k}{3}$ Elemente **then**
    BALANCE p
else
  Gib Fehlermeldung „Schlüsselwert x nicht im Baum.“ aus

Algorithmus 2 BALANCE p

if Page p ist nicht die Wurzel **then**
  if Page p hat einen linken oder rechten Bruder mit mehr als $\frac{4k}{3}$ Elementen (ggf. 2 linke bzw. 2 rechte Brüder betrachten) **then**
    Hole ein Element aus Bruder
  else
    Vereinige Page p mit beiden Brüdern und beiden Trennschlüsseln aus dem Vaters ϕp zu zwei Pages und einem Trennschlüssel im Vater.
    if Unterlauf im Vaterknoten ϕp **then**
      BALANCE ϕp

Dabei ist zu beachten, daß die Wurzel nun bis zu $\frac{8k}{3}$ Schlüsselwerte enthalten darf, um sie bei einem Überlauf in zwei Knoten mit jeweils $\frac{4k}{3}$ Schlüsselwerten und eine neue Wurzel mit einem Schlüsselwert zerteilt zu können.

(b) Bestimmung der maximalen Höhe eines solchen B-Baumes.

Sei n die Anzahl der im Baum T gespeicherten Schlüsselwerte, h die Höhe von T. Dann gilt

$$\begin{aligned} n &\geq 1 + \sum_{i=0}^{h-1} 2\left(\frac{4k}{3}+1\right)^i \cdot \frac{4k}{3} \\ &= 1 + 2\left(\left(\frac{4k}{3}+1\right)^h - 1\right) \\ &= 2\left(\frac{4k}{3}+1\right)^h - 1 \end{aligned}$$

Nach h aufgelöst erhält man also

$$
\begin{aligned}
h &\leq \log_{(\frac{4k}{3}+1)}\left(\frac{n+1}{2}\right) \\
&= \frac{\ln\frac{n+1}{2}}{\ln(\frac{4k}{3}+1)}
\end{aligned}
$$

(c) Algortihmus 3 beschreibt eine weitere Variante der INSERT-Operation, so daß alle Knoten außer der Wurzel mindestens $\frac{3k}{2}$ Schlüsselwerte enthalten.

Algorithmus 3 INSERT x

Suche Element x im Baum
if Suche endet erfolglos im Blatt p **then**
    Füge x in p ein
    if p enthält nun mehr als $2k$ Elemente **then**
        BALANCE' p
else {Element x gefunden}
    Gib Fehlermeldung „Schlüsselwert x bereits vorhanden.“ aus

Algorithmus 4 BALANCE' p

if Page p ist nicht die Wurzel **then**
    if Der linke (bzw. rechte) Bruder βp von p hat weniger als $2k$ Schlüsselwerte **then**
        Verschiebe den Trennschlüssel von p und βp aus dem Vater in den Bruder βp
        Verschiebe den kleinsten (bzw. den größten) Schlüsselwert aus p an die Stelle des Trennschlüssels im Vater von p
    else
        Vereinige Page p mit beiden Brüdern und beiden Trennschlüsseln aus dem Vaters zu vier Pages und drei Trennschlüsseln im Vater.
        if Überlauf im Vaterknoten ϕp **then**
            BALANCE' ϕp
else {Page p ist die Wurzel}
    if Page p enthält mehr als $3k$ Schlüsselwerte **then**
        Splitte p (d.h. die Wurzel) in zwei Knoten mit jeweils $\frac{3k}{2}$ Schlüsselwerten und einen neuen Wurzelknoten mit einem Schlüsselwert.

Beachte: Die Wurzel hier erst gesplittet, wenn sie bereits $3k$ Schlüsselwerte enthält und ein weiterer eingefügt werden soll.