

Übungsblatt 7 – Lösungsvorschläge

für die freiwillige Übung

Unterlagen zur Vorlesung: „www.dvs.informatik.uni-kl.de/courses/DBSREAL/“

Aufgabe 1: Analyse einer Hash-Funktion

Die Schlüssel zur Identifikation der Datensätze einer Anwendung seien als CHAR(8) im EBC-DIC-Code definiert. Sie werden typischerweise als Zahlen in festen Abständen, also 10, 20, 30, ..., vergeben. Vor Verwendung in einer Hash-Funktion werden diese Schlüssel erst nach folgendem Schema gefaltet und dabei mit EXOR verknüpft. In hexadezimaler Darstellung (im EBC-DIC-Code) ergibt sich beispielsweise nach Faltung für den Schlüssel S_1 mit 0000 0010 der Wert von K_1 mit

$$K_1 = \begin{array}{cccccccc} F & 0 & F & 0 & F & 0 & F & 0 \\ F & 0 & F & 0 & F & 1 & F & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0_{16} \end{array} \oplus \triangleq 256_{10}$$

Danach werde das Divisionsrest-Verfahren mit

$$h(K_i) = K_i \bmod n$$

angewendet.

Analysieren Sie das Kollisionsverhalten dieser Funktion bei obiger Schlüsselvergabe und einem Hash-Bereich von $n = 576$ Buckets.

- a) Nach welchen Abständen von j vergebenen Schlüssel, die sich jeweils mit einem Inkrement von Δk (was dem Wert von K_1 entspricht) erhöhen, treten Kollisionen auf?

Kollision, wenn

$$K_i \bmod n = K_j \bmod n = (K_i + t \cdot n) \bmod n, \quad t = 1, 2, 3, \dots$$

Schlüsselvergabe sei $K_j = K_i + j \cdot \Delta k, \quad j = 1, 2, 3, \dots$

Kritisch, wenn

$$j \cdot \Delta k = t \cdot n$$

Nach welchem Abstand $j \cdot \Delta k$ ergibt sich eine Kollision?

$$j = (t \cdot n) / \Delta k, \quad n = 576_{10} = 2^6 \cdot 3^2 \\ \Delta k = 100_{16} = 256_{10} = 2^8$$

$$j = (t \cdot 576) / 256 = (t \cdot 2^6 \cdot 3^2) / 2^8 = (t \cdot 3^2) / 2^4$$

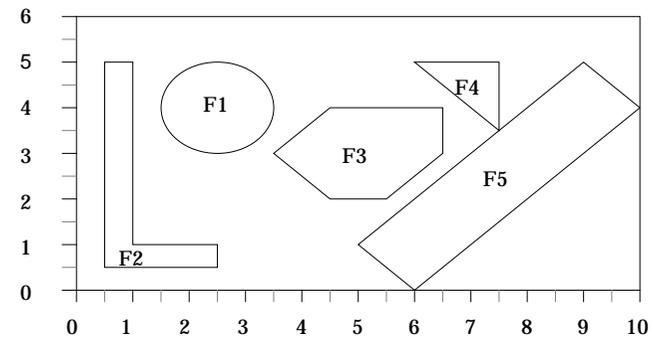
$j = 9$ mit bei einem minimalen $t = 4$, da j ganzzahlig ist.

- b) Was ist eine effektive Abhilfemaßnahme?

Für die Größe des Hash-Bereichs sollte eine Primzahl gewählt werden. Dann ergeben sich keine gemeinsamen Teiler von Schlüsselinkrement Δk und n , so daß die Kollisionsabstände maximal werden.

Aufgabe 2: R-Baum

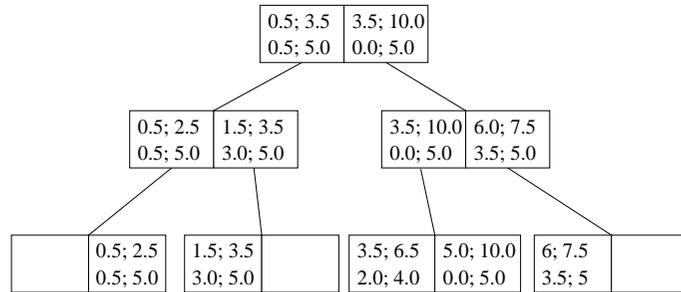
Gegeben ist die folgende Anordnung von 2-dimensionalen Objekten. Speichern Sie diese in der Reihenfolge $F1 - F5$ in einem R-Baum mit $M = 2$ ab.



Aufbau der Einträge: x-koor. unten links x-koor. oben rechts
y-koor. unten links y-koor. oben rechts

1. Entscheidungskriterium: x-koor unten links,
2. Entscheidungskriterium: x-koor. oben rechts

Nach dem Eintragen aller Flächen.



Führen Sie die folgenden Suchoperationen unter Benutzung des erhaltenen R-Baums aus:

a) Finden Sie alle Figuren die den Punkt (6, 3.5) beinhalten.

Die Suche im R-Baum liefert uns als mögliche Kandidaten die Figuren F3 sowie F5. Da aber die umfassenden Rechtecke nicht exakt der Figurkontur entsprechen muß, muß für die gefundenen Figuren einzeln das Enthaltensein des Punktes überprüft werden, so daß letztendlich lediglich F3 übrigbleibt.

b) Welche Figuren sind in dem durch die Punkte (3.5, 0), (8, 5.5) bestimmten Rechteck (Fenster) vollständig enthalten?

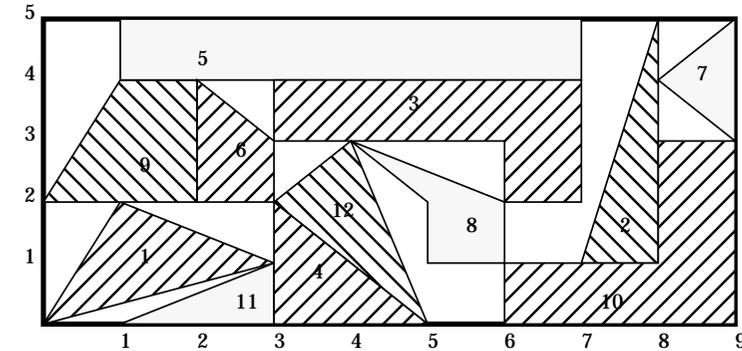
Diese Frage ist unter ausschließlicher Nutzung der R-Baum-Einträge zu beantworten und resultiert in F3, F4

c) Bestimmen Sie alle mit dem Fenster (2, 2), (4, 4) überlappenden Figuren.

Hier liefert uns der R-Baum wiederum nur mögliche Kandidaten, nämlich F1 und F3. Für diese muß wieder konkret überprüft werden, ob sie überhaupt in dem Suchfenster liegen, was hier der Fall ist.

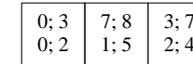
Aufgabe 3: R-Baum - Suchverfahren

Gegeben sei folgende Anordnung von Flächen.

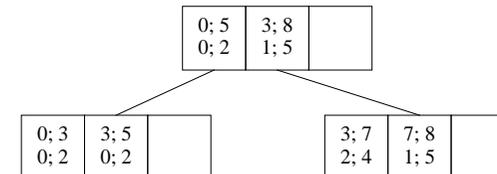


a.) Tragen Sie diese in der durchnummerierten Reihenfolge in einen R-Baum mit den Parametern M=3 (maximale Anzahl von Einträgen in einem Knoten) und m=2 (minimale Anzahl von Einträgen in einem Knoten) ein.

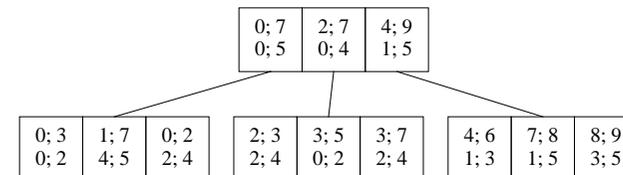
Nach dem Eintragen der Flächen 1, 2, 3



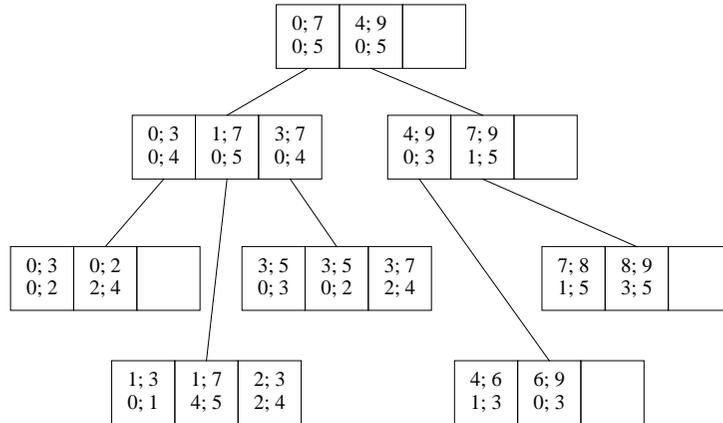
Nach dem Eintragen von Fläche 4



Nach dem Eintragen von Flächen 5, 6, 7, 8, 9
(Beim Eintragen von F8 wird F3 verschoben)



Nach dem Eintragen aller Flächen



b.) Schreiben Sie die Funktionen

FINDE_ALLE_VOLLSTAENDIG_ENTHALTENEN_FLAECHEN (Rechteck)

Funktion braucht nur auf dem R-Baum zu arbeiten, die konkreten Figuren interessieren nicht, da diese vollständig durch die umgebenden Rechtecke beschrieben werden.

FINDE_ALLE_TEILWEISE_ENTHALTENEN_FLAECHEN (Rechteck)

Der R-Baum kann uns hierbei nur die Flächen liefern, die möglicherweise das gegebene Rechteck schneiden. Für jede dieser Flächen ist eine konkrete Überprüfung notwendig!

die alle Flächen findet, die vollständig resp. teilweise in dem gegebenen Rechteck enthalten sind.

c.) Testen Sie die Funktionen mit dem Rechteck, daß durch die Eckpunkte (3,1) und (7,4) beschrieben wird.

Vollständig: 3, 8

Teilweise: 12, 4 (+ vollständig 3, 8)

d.) Welche Verbesserung des R-Baums erlaubt eine einfachere Suche?

Der R*-Baum, da der Suchraum durch überlegte Zerlegung in Teilflächen eingeschränkt wird. (Hier eine Zerlegung bei x=3 und x=7)

Beachte aber größeren Overhead in der Baumorganisation