

Einführung „R-Baum“

Softwarepraktikum „Datenstrukturen“

Sommersemester 2004

RWTH Aachen, Lehrstuhl für Informatik IX

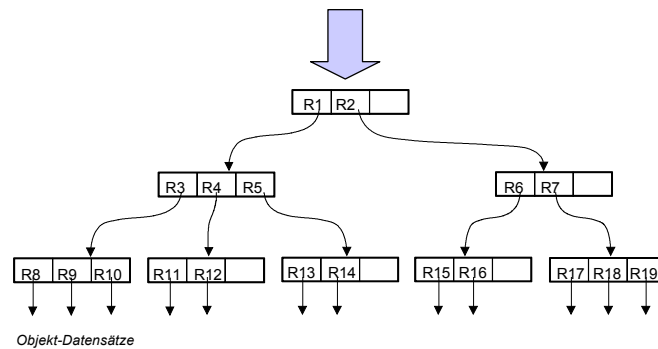
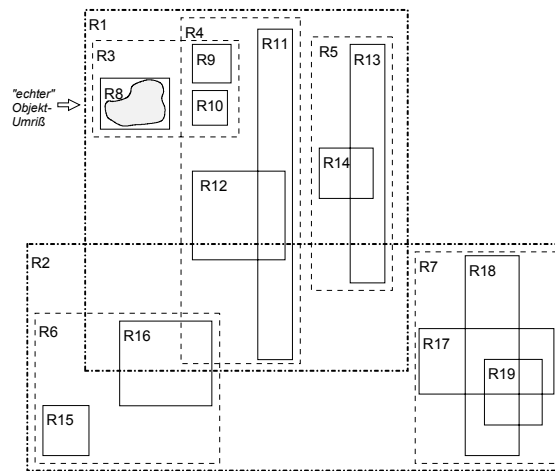
Softwarepraktikum „Datenstrukturen“ - 2

Einführung R-Baum

Allgemeines:

- Datenstruktur für **räumlich ausgedehnte Daten** (hier: 2-D)
- angelehnt an B-Baum
- höhenbalanciert
- seitenorientierte Knotenstruktur
- Unterscheidung:
 - Nichtblatt-Knoten (Suchraum-Partitionierung)
 - Blatt-Knoten (Verweis auf Datenobjekte)
- Approximation der Objekte durch umschließende Rechtecke
 - Blatt-Knoten approximieren Objekt (hier: Objekte selber sind ebenfalls Rechtecke)
 - Nichtblatt-Knoten approximieren Sohnknoten
- Suchraum-Aufteilungen i. a. nicht disjunkt
- Literatur:
 - R-Baum:
<http://www.sai.msu.su/~megeera/postgres/gist/papers/gutman-rtree.pdf>

Beispiel:



Objekt-Datensätze

Eigenschaften des R-Baums:

M : maximale Anzahl von Einträgen pro Knoten,
 $m \leq M/2$: Mindestbelegung pro Knoten

- Anzahl der Index-Einträge pro Blatt-Knoten zwischen m und M
- Anzahl der Sohn-Knoten von Nichtblatt-Knoten zwischen m und M
- in Nichtblatt-Knoten ist das kleinste Rechteck gespeichert, welches Rechtecke der Sohnknoten umfasst
- in Blatt-Knoten ist Objekt und sein kleinstes umschließendes Rechteck gespeichert (hier: Objekt = Rechteck)
- höhenbalanciert (Blattknoten auf derselben Höhe)
- Höhe des Baums $\leq \lceil \log_m N \rceil - 1$

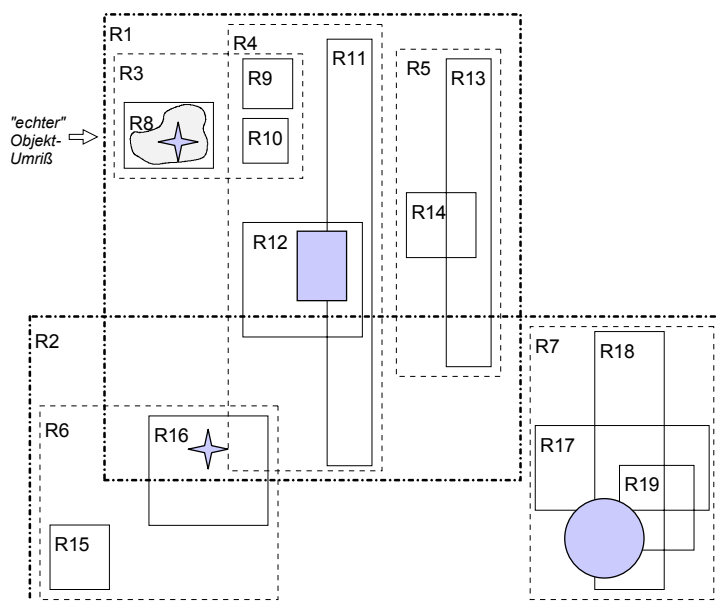
Eigenschaften des R-Baums (2):

M : maximale Anzahl von Einträgen pro Knoten,
 $m \leq M/2$: Mindestbelegung pro Knoten

Motivation für M und m:

- Datenstruktur ist plattenorientiert (Blockstruktur)
- ein Knoten pro Plattenblock
- M maximale Anzahl an Einträgen, die in einen Block/Knoten passen
- Wahl von Blockgröße/M:
 - groß: viele Einträge in Knoten, breiter Baum
 - klein: höherer Baum, langsames Durchlaufen des Baums
- Wahl von m: „performance tuning“

Suchen im Baum



Punktanfrage: 

Rechteckanfrage: 

- Überlappung
- Enthaltensein

Bereichsanfrage: 

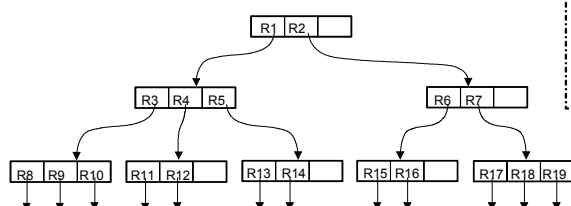
Einfügen neuer Objekte

- Einfügungen stets in den Blatt-Knoten
 - Durchlaufen ab Wurzel
 - jeweils Wahl des Sohnknotens, dessen Rechteck am wenigsten vergrößert werden muss
 - bei Ankunft im Blattknoten: Einfügen des neuen Elements
- bei Überlauf: Split des Blatt-Knotens, Einfügen von Eintrag in Vaterknoten, ggf. Split des Vaterknotens etc.
- Anpassen der Rechtecke der übergeordnete Knoten

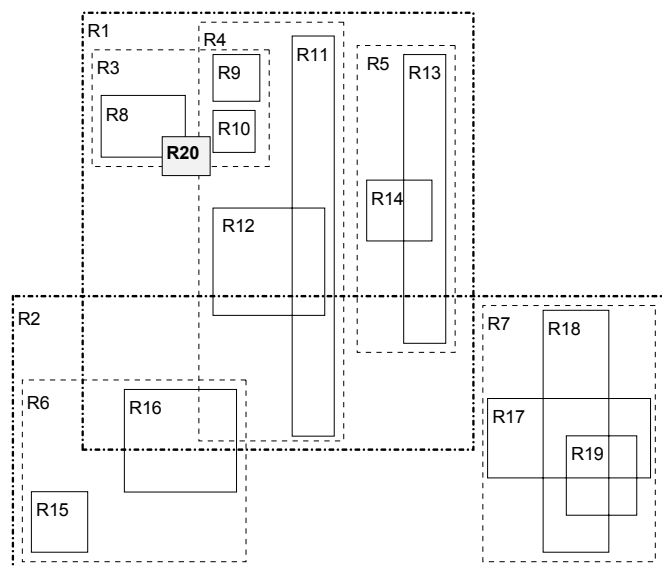
Beispiel „Einfügen von Elementen“

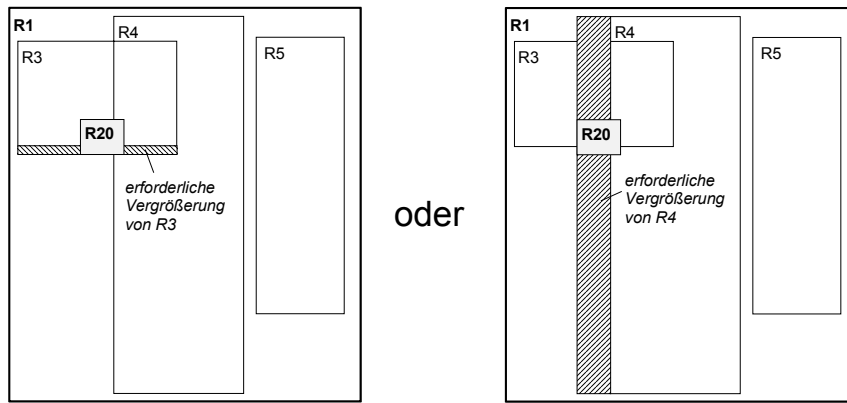
Einfügen von R20:

1. Start bei Wurzel
2. R1 umfasst R20
komplett (im Gegensatz zu R2) → wähle R1
3. untersuche Teilbaum R1
4. Problem: Einfügen in R3 oder R4?

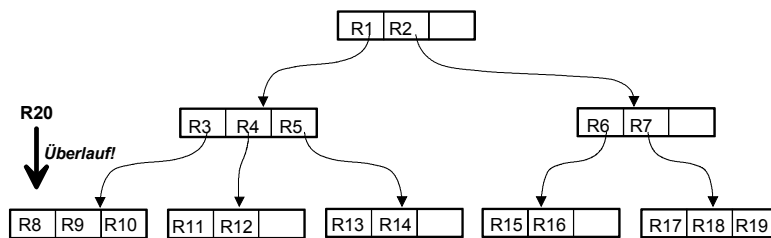


Objekt-Datensätze

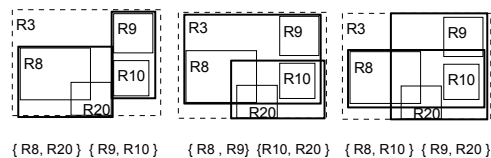
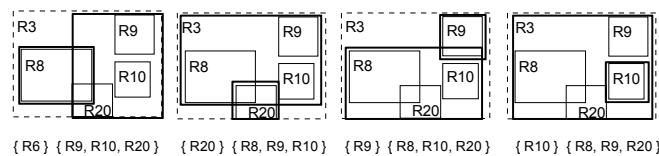
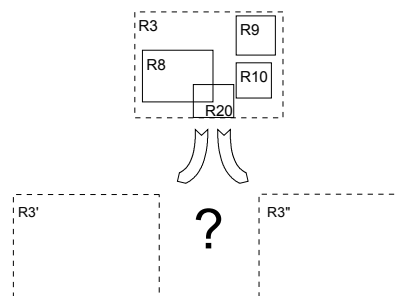




hier: Wahl von R3



Vorgehensweise bei Knotenüberlauf

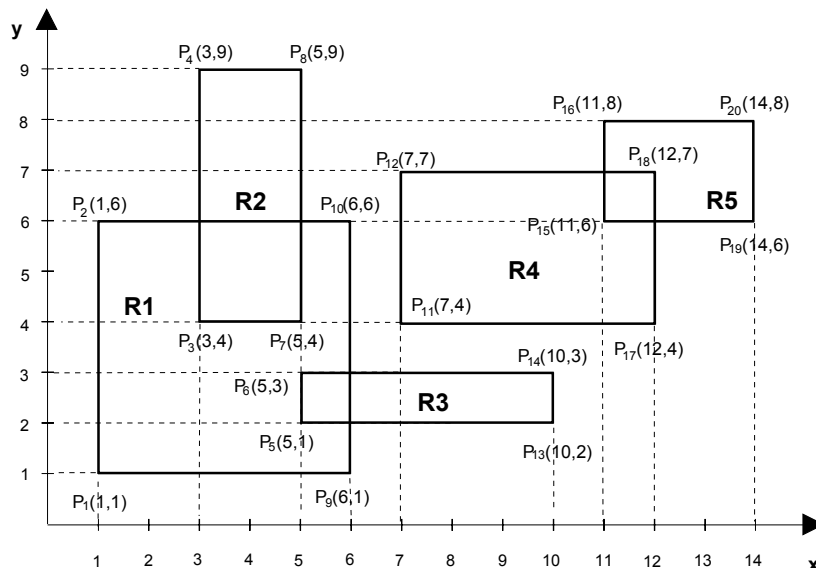


Knotenüberlauf

- große Anzahl Möglichkeiten, Knoten aufzusplitten
- optimale Lösung: Berechnung zu aufwändig
- daher: Heuristiken, z.B. „Quadratic Split“
 - 1) Suche 2 Elemente mit höchstem „Verschnitt“ und verteile diese auf beide Gruppen
 - 2) - alle Elemente verteilt: ENDE
- falls restlichen Elemente benötigt werden, um Mindestbelegung einer der Gruppen zu garantieren: verteile Elemente auf diese Gruppe, ENDE
 - 3) Berechne für jedes Element
 - a) $d1$: Zuwachs des Rechtecks der ersten Gruppe (bei Hinzufügen des Elements)
 - b) $d2$: Zuwachs des Rechtecks der zweiten Gruppe
 - 4) wähle Element mit maximaler Differenz zwischen $d1$ und $d2$
 - 5) fahre fort bei 2)

Berechnung „Verschnitt“

$$d(R1 \oplus R2) = 5 * 8 - \text{fläche}(R1) - \text{fläche}(R2) = 40 - 25 - 10 = 5$$



$$d(R1 \oplus R2) = 5$$

$$d(R1 \oplus R3) = 15$$

$$d(R1 \oplus R4) = 26$$

$$d(R1 \oplus R5) = 60$$

$$d(R2 \oplus R3) = 34$$

$$d(R2 \oplus R4) = 20$$

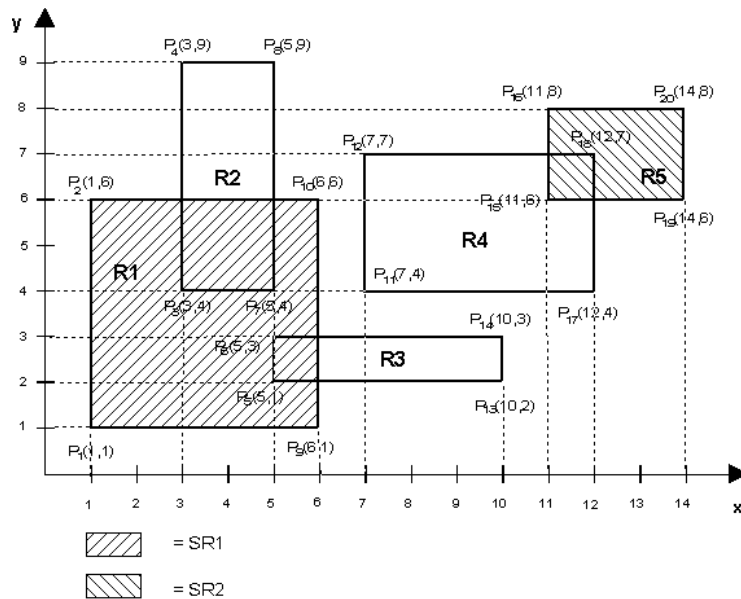
$$d(R2 \oplus R5) = 39$$

$$d(R3 \oplus R4) = 15$$

$$d(R3 \oplus R5) = 43$$

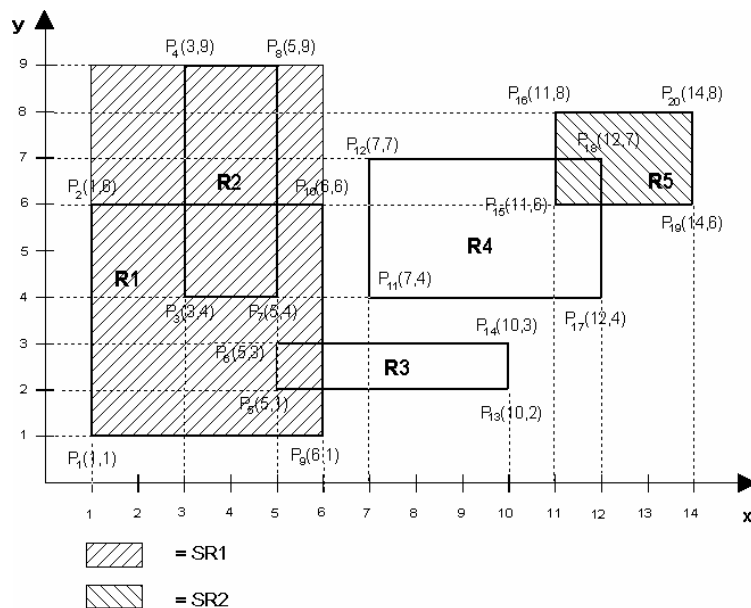
$$d(R4 \oplus R5) = 7$$

Verteilung auf Knoten



falls x =	SR1	SR2	Differenz
$d(R2 \oplus x)$	5	39	34
$d(R3 \oplus x)$	15	43	28
$d(R4 \oplus x)$	26	7	19

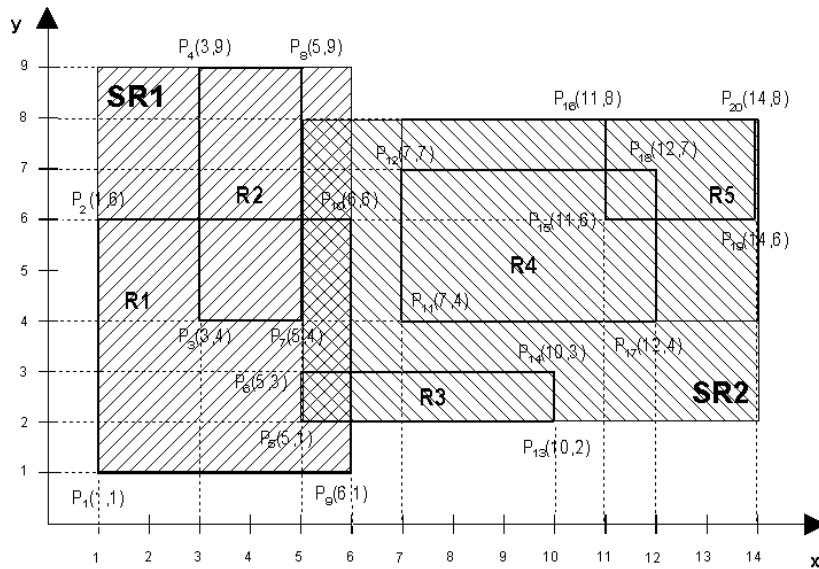
Verteilung auf Knoten (2)



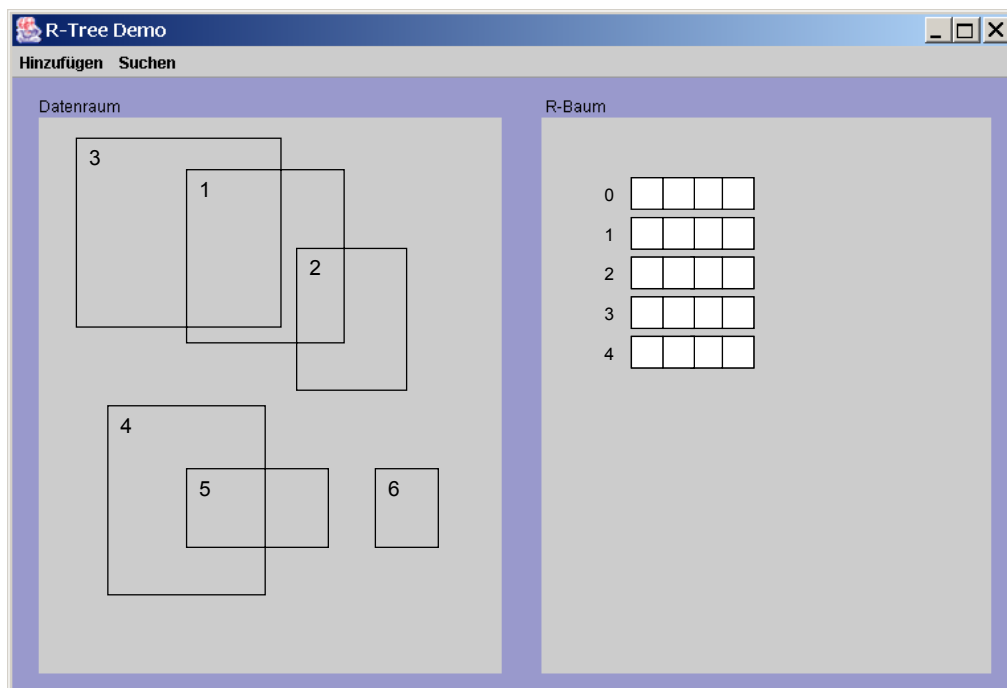
falls x =	SR1	SR2	Differenz
$d(R3 \oplus x)$	27	43	16
$d(R4 \oplus x)$	33	7	26

Verteilung auf Knoten (3)

endgültige Verteilung:



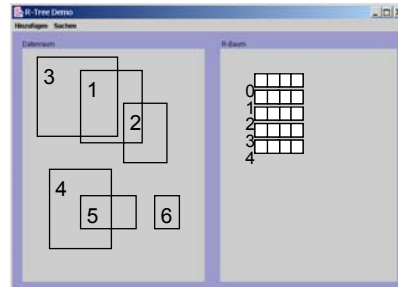
Visualisierung (Beispiel)



Visualisierung (Beispiel)

Anfangsparameter:

- maximale Belegung „M“ (abhängig von Blockgröße)
- minimale Belegung (mittels „m“)



linke Seite:

- Sicht des Datenraums
- Eingabe der Rechtecke mit der Maus
- Darstellung der umgebenden Rechtecke

rechte Seite:

- Sicht der Blockstruktur (Darstellung der einzelnen Plattenblöcke)
- bei Mausklick: „Zoom“ des Blockes und Darstellung der Verweise/Rechtecke