

R-Bäume*

Hans Jürgen Ohlbach

23. März 2018

Keywords: Suchbäume, R-Bäume

Empfohlene Vorkenntnisse: Bäume, siehe Baeume.pdf, B-Bäume, siehe BBaeume.pdf

Inhaltsverzeichnis

1	Motivation	2
2	Die Struktur von R-Bäumen: Eindimensional	3
3	R-Bäume: Zweidimensional	4
4	R-Bäume: Drei- und Höherdimensional	5

*Dieser Text ist Teil einer Sammlung von Miniskripten zur Einführung in die Informatik. Er ist in erster Linie für Nichtinformatiker gedacht, kann aber natürlich auch als erste Einführung für Informatiker nützlich sein.

1 Motivation

Man stelle sich eine Geodatenbank vor, in der alle Orte, Straßen und sogar Häuser Deutschlands zusammen mit ihren geographischen Koordinaten verzeichnet sind. Ein GPS-Gerät misst die Koordinaten meines Standorts. Jetzt möchte ich herausfinden, in welchem Ort, welcher Straße oder sogar welchem Haus ich bin. Die einfachste Vorgehensweise wäre, die Liste aller Orte, Straßen und Häuser durchzusuchen, und jeweils deren Koordinaten mit den Koordinaten des GPS-Geräts zu vergleichen. Das kann dauern!

Eine erste Verbesserung wäre, Deutschland in z.B. Quadratkilometer große Quadrate einzuteilen, und in der Geodatenbank jedem dieser Quadrate die Orte, Straßen und Häuser darin zuzuordnen. Jetzt würde man als ersten Schritt aus den Koordinaten des GPS-Geräts das Quadrat ermitteln, und dann nur die Objekte in dem Quadrat durchsuchen. Das geht schon bedeutend schneller.

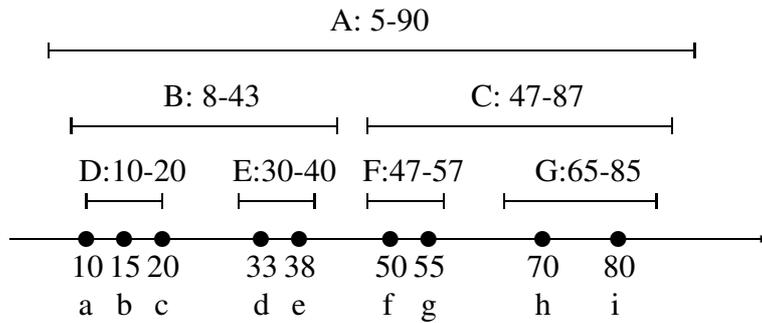
Es hat aber auch einen Nachteil: es gibt nämlich Quadrate, in denen ganz viele Objekte sind, und Quadrate, in denen gar keine Objekte sind. Das ist keine befriedigende Lösung.

Die Idee mit den Quadraten ist im Prinzip nicht schlecht. Man müsste sie nur flexibler strukturieren: in Ballungsgebieten kleiner, und auf dem Land größer. Idealerweise wäre die Aufteilung so, dass in jedem Quadrat gleich viele Objekte sind.

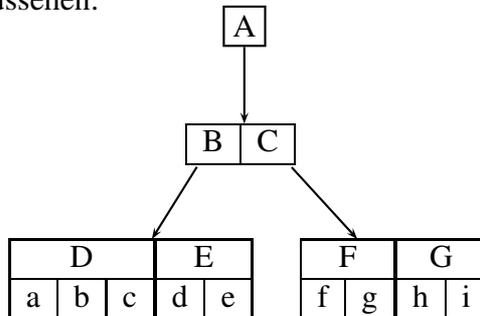
Genau diese Idee realisieren R-Bäume. Die Knoten entsprechen geometrischen Regionen. Es müssen dabei aber nicht unbedingt Quadrate sein. Die Hierarchie im Baum wird genutzt, um große Regionen in immer kleinere Regionen aufzuteilen. An den Blattknoten stehen dann die kleinsten Regionen, mit den Verweisen auf die Objekte, die in diesen Regionen sind.

2 Die Struktur von R-Bäumen: Eindimensional

R-Bäume sind im Prinzip für beliebige Dimensionen geeignet. Wir illustrieren sie daher zunächst für eine Dimension. Eine mögliche Anwendung wäre ein Straßenzug mit Häusern. In der Graphik unten sei ein Straßenzug mit Häusern a-i, zusammen mit ihren Koordinaten dargestellt. Der Straßenzug wird in Intervalle eingeteilt. Das Intervall A umfasst den ganzen Straßenzug. A enthält die Intervalle B und C. B enthält E,F, und C enthält G und H.

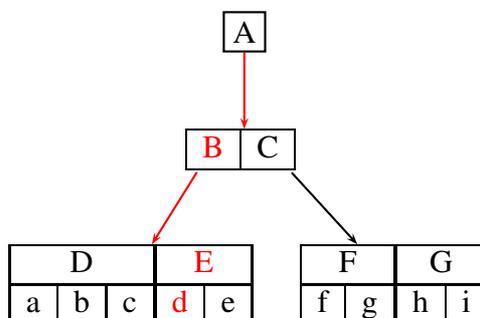


Ein R-Baum dafür könnte so aussehen:



Natürlich sind mit den einzelnen Intervall- und Punktnamen auch deren Koordinaten assoziiert.

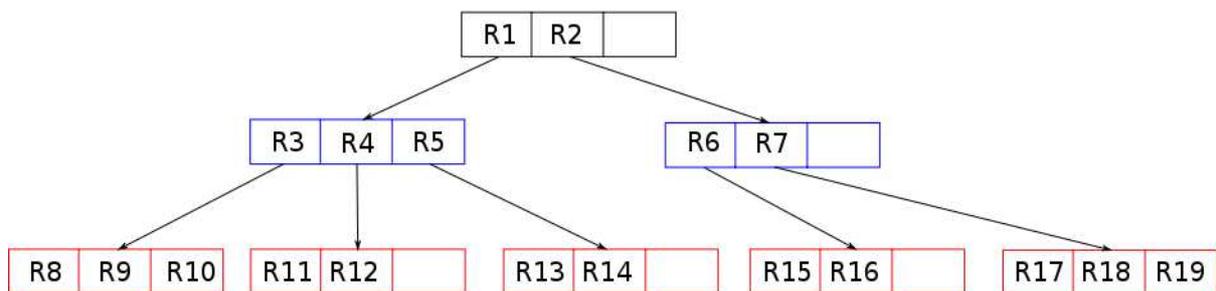
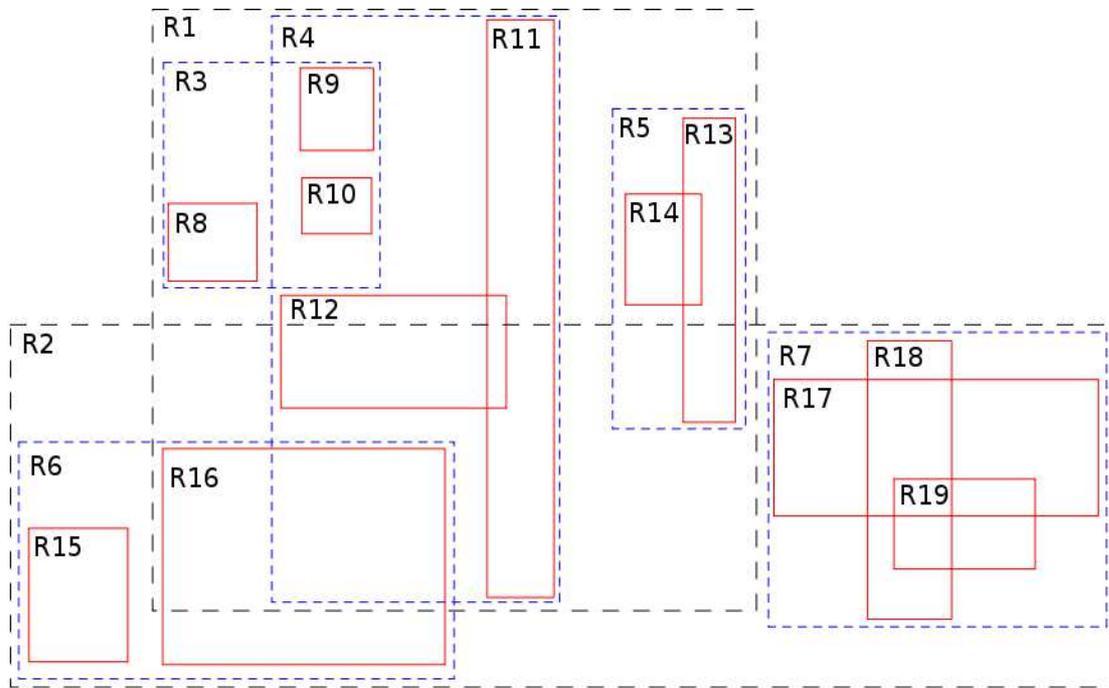
Angenommen ich suche jetzt das Haus mit den Koordinaten 33. 33 liegt im Intervall A:5-90. Eine Ebene tiefer stellt man fest, dass es im Intervall B:8-43 liegt. Das Intervall C ist damit ausgeschlossen. Eine weitere Ebene tiefer kommt man zu E:30-40. In E liegt das Haus d mit den Koordinaten 33. Der Suchpfad (rot) führt also direkt zu d.



Sucht man stattdessen nach der Koordinate 35, landet man auch in E, und kann dann feststellen, dass der Punkt zwischen den Häusern d und e liegt.

3 R-Bäume: Zweidimensional

Das nächste Beispiel zeigt ein zweidimensionales Szenario.¹ Die Ebene ist dabei in Rechtecke aufgeteilt, die sich auch überlappen können. Der R-Baum darunter ist dabei so strukturiert, dass jeder Knoten Platz für drei Regionen hat, wobei nicht alle Plätze belegt sein müssen.



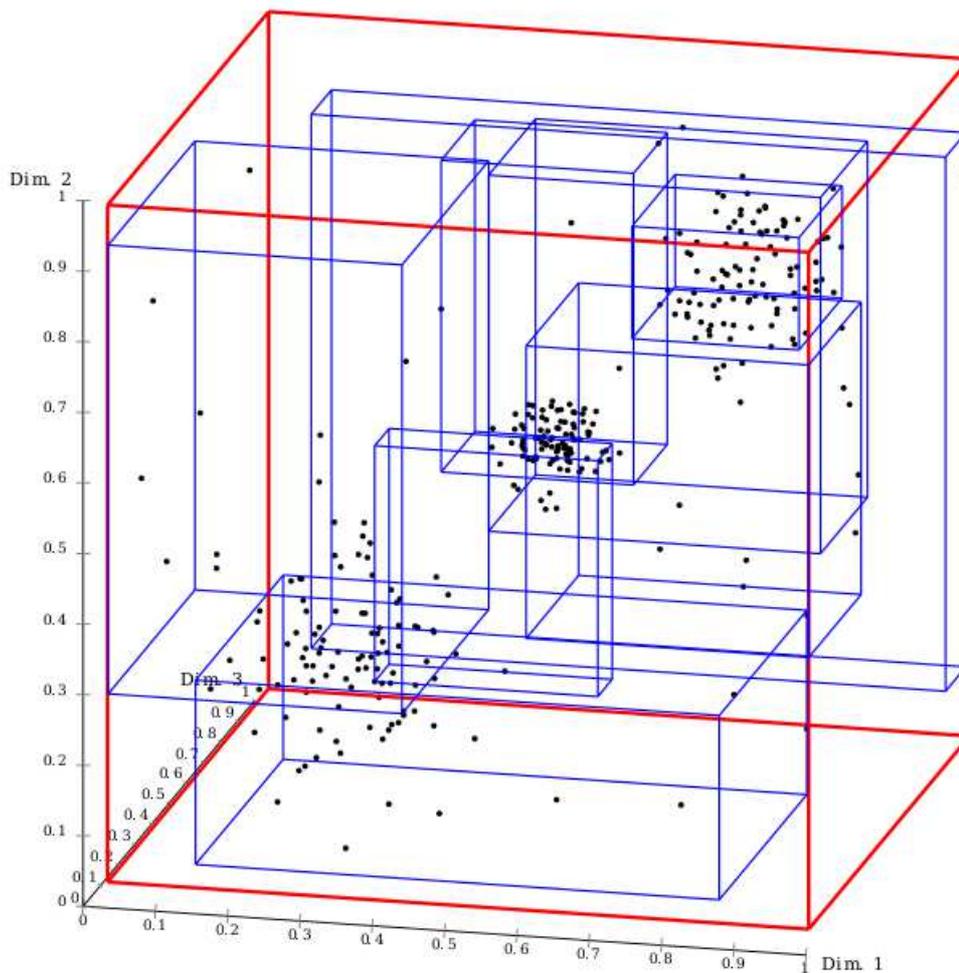
Mit jedem Namen eines Rechtecks sind natürlich die Koordinaten seiner Eckpunkte assoziiert. Mit den Rechtecken in den Blattknoten wären dann die Objekte assoziiert, die in dem Rechteck liegen. Damit kann man z.B. ein Objekt in R13 suchen, indem man zunächst feststellt, dass es in R1 liegt, darin in R5, und darin in R13.

Das Beispiel zeigt auch Überlappungen der Rechtecke. Das behindert die Suche nach einem konkreten Punkt nicht, da es im Prinzip egal ist, welchem Pfad man folgt, wenn der Punkt in einer Überlappung liegt.

¹Von User Sinkie für Wikipedia erzeugt.

4 R-Bäume: Drei- und Höherdimensional

Das nächste Bild zeigt eine dreidimensionale Punktwolke.² Der Raum ist dabei in Würfel aufgeteilt, so dass man wiederum einen R-Baum bilden könnte. Der wird allerdings zu komplex und ist hier nicht mehr gezeigt.



Die generelle Idee sollte jetzt klar sein: Man unterteilt einen Raum in immer kleinere Räume, die man dann ganz ähnlich wie bei B-Bäumen den Knoten zuteilt. Die Räume auf einer Ebene des R-Baumes müssen nicht unbedingt disjunkt sein.

Auf der Suche nach einem konkreten Punkt testet man für einen Knoten, in welcher Region er liegt, und folgt dann dem Verweis zu dem entsprechenden Kindknoten. An dem Blattknoten angekommen, findet man die Objekte, die in der Region liegen. Da in einer Region mehrere Objekte liegen können, kann man auch Anfragen beantworten der Art: welche Objekte liegen in der Nähe einer gegebenen Koordinate. (sog. *Nächste-Nachbarn-Anfragen*).

²Das Bild wurde mit dem ELKI-Tool des Instituts für Informatik der LMU erzeugt.

R-Bäume zeigen ausreichend gute Performanz bis zur Dimension 10. Für höhere Dimensionen, wie sie z.B. in der Bioinformatik vorkommen, wurden Verfeinerungen der R-Bäume entwickelt: R*-Bäume, R+-Bäume, X-Bäume, M-Bäume usw. Insbesondere bei R*- und R+-Bäumen werden Überlappungen reduziert, oder ganz vermieden, so dass die Aufteilung in Teilregionen verfeinert wird, und dadurch die Suche nach einem konkreten Objekt schneller wird.

Die Algorithmen für den Aufbau solcher Bäume, das Einfügen und Löschen von Knoten sind komplex. Für Interessierte wird daher auf die Originalliteratur verwiesen.

Weitere Baumtypen für spezielle Anwendungen sind Quadtree, K-d-Baum, UB-Baum, Bereichsbaum, Gridfile.