

Neue Trends zur Suche in modernen Datenbanksystemen
Wintersemester 2013/14

Übungsblatt 2: Distanzen / R-Bäume / Indexbasierte Anfragen

Besprechung: 06.11.2013

Aufgabe 2-1 MINDIST / MAXDIST

Berechnen und zeichnen Sie für $i = 1, 2$ die folgenden Distanzen zwischen den Anfragepunkten Q_i und dem Minimum Bounding Rectangle R wie dargestellt in Abbildung 1: $\text{MINDIST}(Q_i, R)$ und $\text{MAXDIST}(Q_i, R)$. Zugrundeliegende Metrik ist die durch die euklidische Norm induzierte Metrik. Der Abstand zweier Linien des Gitternetzes betrage 1cm.

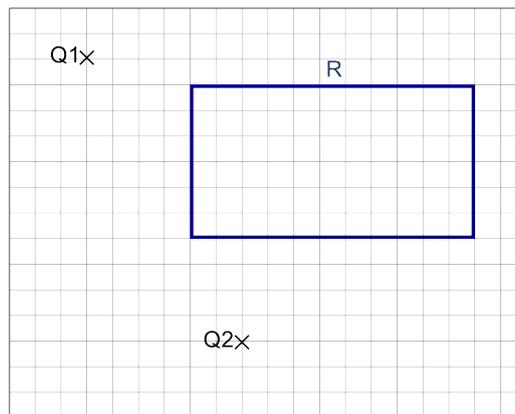


Abbildung 1: MBR R und Anfragepunkte Q_1, Q_2 .

Aufgabe 2-2 MINDIST / MAXDIST zwischen zwei MBRs

Berechnen und zeichnen Sie für $i = 1, 2$ die folgenden Distanzen zwischen den MBRs R_i und dem MBR R wie dargestellt in Abbildung 2: $\text{MINDIST}(Q_i, R)$ und $\text{MAXDIST}(Q_i, R)$. Zugrundeliegende Metrik ist die durch die euklidische Norm induzierte Metrik. Der Abstand zweier Linien des Gitternetzes betrage 1cm.

Aufgabe 2-3 L_p -MINDIST

Man gebe die Berechnung der MINDIST-Funktion zwischen einem Anfragepunkt und einer rechteckigen Seitenregion (achsenparallel) in einem d -dimensionalen Datenraum für folgende Metriken an:

- (a) Manhattan-Distanz (L_1)
- (b) Maximums-Distanz (L_∞)
- (c) gewichtete Euklidische Distanz.

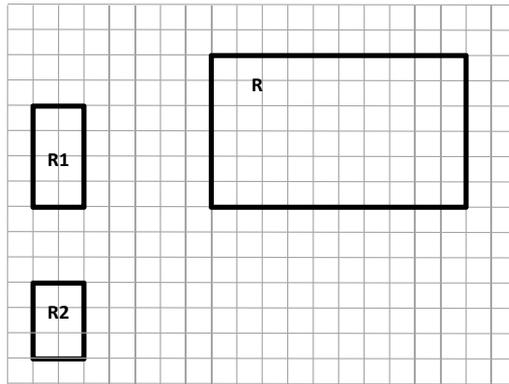


Abbildung 2: Drei Rechtecke

Aufgabe 2-4 R-Bäume / Indexbasierte Anfragen

(a) Fügen Sie in den in Abbildung 2 dargestellten R-Baum die folgenden Datenpunkte ein (Einfügestrategie geringster Flächenzuwachs): $G = (14, 3)$, $H = (7, 11)$, $I = (9, 12)$. Die maximale Seitenkapazität sei $M = 3$, die minimale Anzahl an Einträgen pro Seite (außer der Wurzel) sei $m = 2$. Verwenden Sie als Splitstrategie den *quadratischen Algorithmus*. Dieser ist definiert wie folgt:

- Wähle das Paar von Rechtecken R_1 und R_2 mit dem größten Wert für “toten (leeren) Raum”, im MUR, falls R_1 und R_2 in den selben Knoten kämen.
- Setze $K_1 := \{R_1\}$ und $K_2 := \{R_2\}$
- Solange noch unzugeleitete Knoten vorhanden:
 - Wenn alle restlichen R_i benötigt werden, um den kleineren Knoten minimal zu füllen: teile sie alle dem kleineren Knoten zu.
 - Sonst: Wähle ein unzugeleitetes R_i und teile es dem Knoten zu, der den kleineren Flächenzuwachs erfährt.

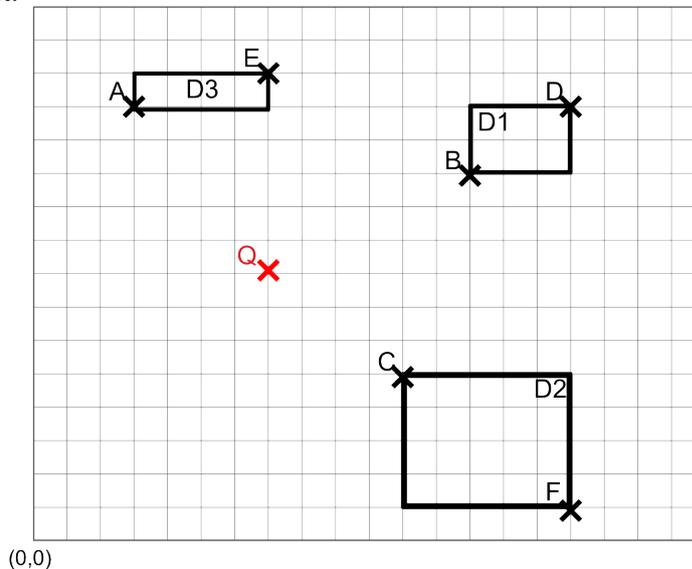


Abbildung 3: Illustration eines R-Baums.

(b) Geben Sie die Resultatmenge einer Indexbasierten Bereichsanfrage (Algorithmus Seite 62 im Skript) um Q mit $\varepsilon = 8$ und der von der Manhattan-Norm induzierten Metrik $dist(x, y) = d_{L_1}(x, y) = \sum_{i=1}^d |x_i - y_i|$ (dabei sind $x, y \in \mathbb{R}^d$ Vektoren) an.