

Prof. Dr. Hans-Peter Kriegel
Dr. Matthias Renz

Übungen zur Vorlesung
Index- und Speicherungsstrukturen für Datenbanksysteme
Wintersemester 2009/10

Blatt 11

Besprechung der Aufgaben: Mittwoch, 20.01.2010

Aufgabe 11.1: Eine wichtige Familie von Metriken wird von den sogenannten L_p -Distanzen gebildet. Für $x, y \in \mathbb{R}^n$ ist eine L_p -Distanz definiert wie folgt:

$$L_p(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}}$$

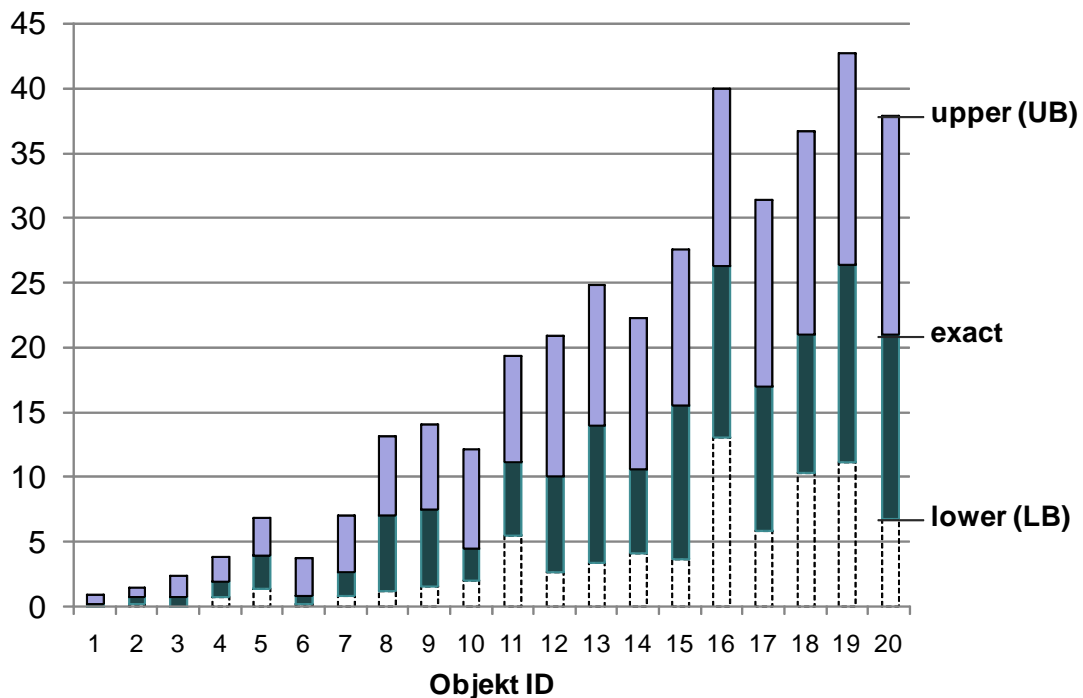
In der Vorlesung wurden Ihnen basierend auf der Euklidischen Distanz (L_2) die Berechnung der minimalen Distanz zwischen einem Punkt q und einer (rechteckigen achsenparallelen) Seitenregion r MINDIST(q,r) vorgestellt (siehe Skript S.123).

- Geben Sie eine entsprechende MINDIST(q,r)-Funktion für die gewichtete Euklidische Distanz an.
- Geben Sie eine entsprechende MINDIST(q,r)-Funktion allgemein für L_p -Normen an.
- Kann in einer ähnlichen Weise auch eine MINDIST(q,r)-Funktion für die quadratische Form angegeben werden? (Begründung durch graphische Erläuterung)
- Neben der MINDIST(q,r)-Funktion, die eine untere Distanz-Schranke zwischen einem Anfragepunkt q und allen Punkten in einer (rechteckigen achsenparallelen) Seitenregion r bildet, ist ebenso die MAXDIST(q,r)-Funktion von großer Bedeutung. Die MAXDIST(q,r)-Funktion berechnet die (kleinste) maximale Distanz zwischen q und allen Punkten in r . Diese wird für die Berechnung der oberen Distanz-Schranke zwischen einem Anfragepunkt q und allen Punkten in einer (rechteckigen achsenparallelen) Seitenregion r verwendet. Geben Sie, analog zur Definition der MINDIST(q,r)-Funktion, die Definition der MAXDIST(q,r)-Funktion auf Basis der allgemeinen L_p -Norm an.

Aufgabe 11.2: Gegeben sei eine Menge von 20 Objekten. Es wird eine mehrstufige Bereichsanfrage ($\epsilon = 5.0$) auf die Objektmenge gestartet. Die entsprechenden Distanzen (Filterdistanz und exakte Distanz) der Objekte zum Anfrageobjekt Q sind in dem unten angegebenen Diagramm dargestellt. Die jeweiligen Objekt-Ids sind auf der X-Achse und die zugehörigen Distanzwerte auf der Y-Achse angegeben. Die Filterdistanz lower entspricht dabei der Distanzabschätzung, die die Untere-Schranken-Eigenschaft (lower bounding) besitzt, während die Filterdistanz upper der Distanzabschätzung entspricht, die die Obere-Schranken-Eigenschaft (upper bounding) besitzt. Die exakten Distanzen exact der Objekte zum Anfrageobjekt Q sind im Filterschritt nicht verfügbar und können erst nach einer Distanz-Verfeinerung (Verfeinerungsschritt) verwendet werden.

Übungen zur Vorlesung
Index- und Speicherungsstrukturen für Datenbanksysteme
Wintersemester 2009/10

Distanz zum Anfrageobjekt

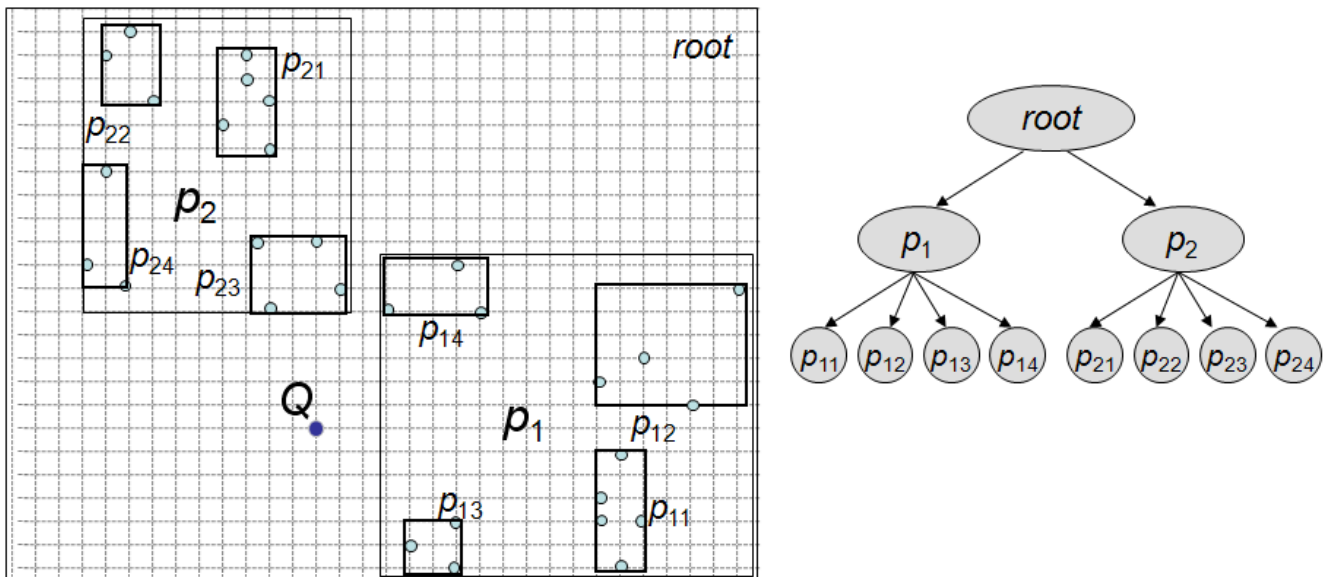


Für die mehrstufige Bereichsanfrage (RQ-MultiStep(DB,Q, ϵ) siehe Skript S. 123) mit $\epsilon = 5.0$ soll aus dem obigen Diagramm für die folgenden Fälle ermittelt werden, welche Objekte einem **Verfeinerungsschritt** unterworfen werden müssen und welche Objekte zum **Resultat** gehören:

- Zum Resultat gehören die Objekte:
- Im Filterschritt ist nur die upper bounding Filterdistanz *upper* verfügbar.
- Im Filterschritt ist nur die lower bounding Filterdistanz *lower* verfügbar.
- Im Filterschritt sind beide Filterdistanzen *lower* und *upper* verfügbar.

Übungen zur Vorlesung
Index- und Speicherungsstrukturen für Datenbanksysteme
 Wintersemester 2009/10

Aufgabe 9.3:



Gegeben sei ein R-Baum der Höhe 3, wie oben skizziert (links: geometrische Darstellung, rechts: schematische Darstellung). Ferner sei ein Anfragepunkt Q gegeben.

- a) Geben Sie im Folgenden die Seiten in der Reihenfolge an in der bei einer Nächste-Nachbar Anfrage $NN(Q)$ unter Verwendung des Algorithmus der **Prioritätssuche** auf sie zugegriffen wird. Als Distanzmaß soll die **Manhattan Distanz (L_1)** verwendet werden.