

Spatial, Temporal and Multimedia Databases I
SoSe 2011

Übungsblatt 2: L_p -Distanz, Mehrstufige Anfragebearbeitung, Bereichsanfrage mit Index, L_p -MINDIST

Besprechung: 16.05.2011

Aufgabe 2-1 L_p -Distanz, Mehrstufige Anfragebearbeitung

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 die Euklidische Distanz (L_2), die Manhattan-Distanz (L_1) und die Maximums-Distanz (L_∞) vorgestellt. In welchem Verhältnis stehen diese drei Distanzmaße zueinander? Was bedeutet das im Kontext der Mehrstufigen Anfragebearbeitung?

Hinweis: Im allgemeinen verwendet man $p \in \mathbb{N}^+ \cup \infty$ (ohne 0). Sogenannte "fractional L_p norms" mit $0 < p < 1$ sind keine Metriken, werden aber in Sonderfällen (hochdimensionale Daten) gelegentlich verwendet.

Aufgabe 2-2 Bereichsanfrage mit Index

Erweitern Sie den Indexbasierten Algorithmus (Tiefensuche) zur Rangequery so, dass bereits ganze Teilbäume von Directory-Seiten möglichst früh als "true hits" erkannt werden. Welches zusätzliche Hilfsmittel aus der Vorlesung benötigen Sie dazu? Diskutieren Sie Vor- und Nachteile Ihrer Lösung.

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.

Ist dieselbe Vorgehensweise einer dimensionsweisen Aufsummierung auch für positiv definite quadratische Formen möglich?