

Spatial, Temporal and Multimedia Databases
SoSe 2012

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

Besprechung: 07./10.05.2012

Aufgabe 2-1 *L_p -Distanz, Mehrstufige Anfragebearbeitung*

Eine wichtige Familie von Normen wird von den sogenannten p -Normen gebildet, die die L_p -Distanzen induzieren.

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 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, Skript S. 56) 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 Vorteile und Nachteile Ihrer Lösung.

Aufgabe 2-3 *L_p -MINDIST*

Geben Sie die 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?