

Neue Trends zur Suche in modernen Datenbanksystemen
Wintersemester 2013/14

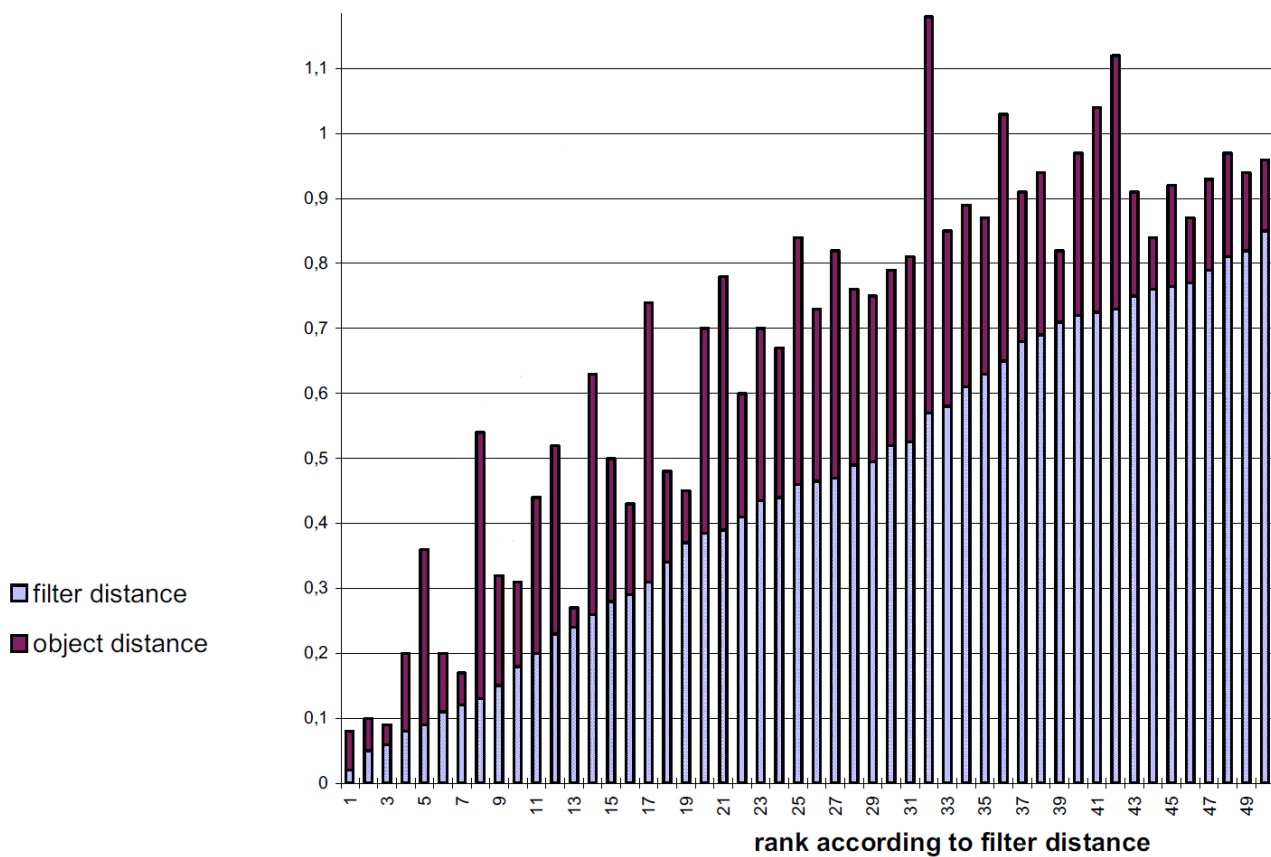
Übungsblatt 4: Multistep

Besprechung: 04.12.2013

Aufgabe 4-1 Mehrstufige Anfragebearbeitung

In der Grafik sind sowohl die im Verfeinerungsschritt ermittelten Objektdistanzen als auch die im Filterschritt ermittelten Featuredistanzen für eine k -NN-Anfrage mit $k=10$ eingezeichnet. Man ermittle, wie viele Objekte bei einer Auswertung nach Priorität einem Verfeinerungsschritt unterworfen werden.

object and
filter distances

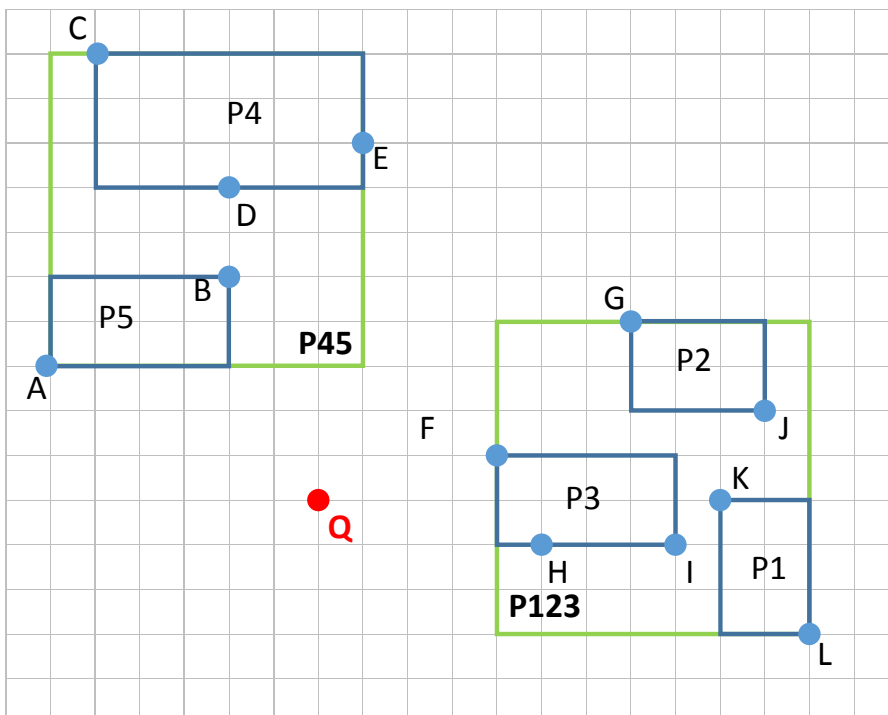


Aufgabe 4-2 Mehrstufige Anfragebearbeitung mit Index

Für Ähnlichkeitsanfragen auf hochdimensionalen Daten (z.B. Zeitreihen) gibt es folgende Technik die unter gewissen Voraussetzungen enorme effizienzgewinne mit sich bringt:

- (a) Überführe die Daten in einen niedrigdimensionaleren Raum (z.B. mittels PCA) in dem die Distanz zwischen zwei Punkte immer eine untere Abschätzung der echten Distanz liefert.
- (b) Indexiere die niedrigdimensionale Darstellung und führe Anfragen mit einem Multistep Algorithmus durch

Gegeben sei also der folgende Index auf den dimensionsreduzierten Daten. Führen Sie eine nächste Nachbar Anfrage unter Verwendung dieser Indexstruktur durch. Die exakten Distanzen der Objekte befinden sich in der nebenstehenden Tabelle.



point	dist to Q
A	10
B	9
C	17
D	12
E	25
F	11
G	14
H	8
I	10
J	16
K	9
L	14