



Wir wünschen frohe Weihnachten und einen guten Rutsch.

Aufgabe 1: Top-k Algorithmen (1 P.)

- (a) Nachfolgend sehen Sie für den NRA-Algorithmus, welche Objekte aus drei Listen sequentiell gelesen wurden, und für den TA-Algorithmus, welche Objekte aus drei Listen sequentiell und leicht abgesetzt welche ungeordnet gelesen wurden. Geben Sie jeweils an, ob der Algorithmus bereits die Top-3-Objekte zurückgeben kann, oder ob er noch weitere Objekte holen muss.

NRA-1

Id	Score	Id	Score	Id	Score
A	7	B	8	C	6
B	3	A	7	D	5
C	2	C	2	B	5

NRA-2

Id	Score	Id	Score	Id	Score
D	11	F	18	L	19
A	10	B	14	B	17
G	5	C	13	A	14

TA-1

Id	Score	Id	Score	Id	Score
G	18	C	27	F	25
F	18	E	22	C	21
A	16	G	19	E	18
C	5	A	15	A	3
E	4	F	12	G	2

TA-2

Id	Score	Id	Score	Id	Score
G	20	C	17	F	19
F	16	E	14	C	18
A	15	G	14	E	13
C	12	A	13	A	10
E	10	F	13	G	9

- (b) Gegeben folgender Top-k-Algorithmus:

1. Wähle zufällig eine der Listen aus
2. Hole die obersten k Einträge, $s := \text{Score}$ des k -ten Eintrags
3. Hole aus allen anderen Listen alle Einträge mit $\text{Score} \geq s$
4. Aggregiere alle geholten Scores und gib die Top- k zurück

Ist dieser Algorithmus korrekt?

¹Quelle: openclipart.org

Aufgabe 2: Skyline Anfragen

(1 P.)

(a) Gegeben ist die folgende Tabelle mit Handytarifen verschiedener Provider:

ID	preis	freiminuten	freisms	netz	handy
1	9.95	100	50	Vodafone	Android
2	4.95	100	100	E-Plus	Android
3	9.95	100	100	Telekom	Android
4	19.95	200	100	O2	iPhone
5	14.95	100	50	Vodafone	Android
6	14.95	50	50	E-Plus	Android
7	18.95	200	100	Telekom	Android
8	18.95	200	100	Telekom	iPhone
9	12.95	100	50	Telekom	Android
10	12.95	0	200	Vodafone	Android

Gesucht ist die Skyline über die folgenden fünf Dimensionen:

- *preis*: billiger ist besser,
- *freiminuten*, *freisms*: mehr ist besser,
- *netz*: Telekom ist besser als Vodafone, Vodafone besser als O2; O2 und E-Plus sind gleich gut,
- *handy*: nicht vergleichbar.

Auf der Vorlesungs-Webseite finden Sie die SQL-Statements zur Erzeugung der Tabelle *handytarif* sowie der Funktion *schlechter*(*netz*, *netz*), die **true** liefert, wenn das erste Netz schlechter als das zweite ist.

Erstellen Sie eine SQL-Anfrage zur Bestimmung der Skyline (ohne den SKYLINE-Operator zu benutzen).

(b) Betrachten wir den Fall, dass bzgl. Dominanz “kleinere Werte” besser sind als größere Werte und ferner Werte aus \mathbb{R}^+ . Zeigen Sie, dass der Nächste Nachbar (NN) zum Ursprung (0,0) auch immer in der Skyline ist.

Aufgabe 3: Erweiterbares Hashing

(1 P.)

Gegeben sei eine Hash-Funktion h , die für die folgenden Schlüssel folgende Werte liefert:

$$h(A) = 00001100 \quad h(B) = 01001100 \quad h(C) = 11101100 \quad h(D) = 10001100 \quad h(E) = 10011100$$

Verteilen Sie unter Verwendung von erweiterbarem Hashing die Datensätze mit den Schlüsseln A, B, C, D und E (in dieser Reihenfolge) in Buckets der Kapazität 2.

Bauen Sie das Directory anhand des Präfixes der Hash-Werte auf. D.h., Sie betrachten die *ersten* d Stellen der Hash-Werte. Da die Buckets zu Beginn leer sind, hat auch das Directory zu Beginn die Größe $d = 0$.

Zeichnen Sie das Directory, sowie die Bucketbelegung nach dem Einfügen der fünf Datensätze. Was ist nun die globale Tiefe d , was sind jeweils die lokalen Tiefen c ?