

Voor preprocessing:

db1: Voor elk numeriek datapunt, maak een gesorteerde lijst van de waarden

db2: Voor elke waarde van elk categorische datapunt, maak een lijst for alle autos gesorteerd op $W(t_u) / (W(t) + W(u))$

db3: Voor elke waarde van elk categorische datapunt, vindt $\log(QF(\text{alles})/QF(\text{dit}))$

db4: Voor elke auto, kijk op hoeveel van de workload queries de auto applied, sla een lookup van auto naar dit op.

Voor querying:

1. Voor categorische data, lees van db.2 de lijst, en lees op volgorde.
2. Voor numerieke data, binSearch in db1 naar de waarde, en lees naar voor en achter.
3. Doe topK op deze lijsten. Met doel functie
4. $1 * \text{numerieke data} + \text{db3}(\text{dit}) * \text{categorische data}$
5. Break ties met db4
6. Break further ties randomly

EXAMPLE

Initial DB:	Searches:
id, brand, price, speed	Ford (3 Times)
0, Ford, 50, 140	fer
1, Ford, 80, 100	vw
2, vw, 2, 30	Ford or vw
3, fer, 10000, 200	Ford or vw
	*(fer or vw or Ford)

Our dbs:

Table: Price

PrimKey SortID, Car

(0 2 2)

(1 0 50)

(2 1 80)

(3 3 10000)

table QF

primKey Brand, QF

Ford 0.5

vw 0.25

fer 0.25

Table: Speed

PrimKey SortID, CarID, Speed

(0 2 30)

(1 1 100)

(2 0 140)

(3 3 200)

Table workloadF

primKey CarID, Workload Frequency

0: 6

1: 6

2: 4

3: 2

table Ford

PrimKey SortID, CarID, Similarity

(0 0 1)

(1 1 1)

(2 2 3/8)

(3 3 1/8)

vw (0 2 1) (1 0 2/7) (2 1 3/8) (3 3 1/5) #same as Ford table

fer (0 3 1) (1 2 1/5) (2 1 1/8) (3 0 1/8) #same as Ford table