Intro To Databases (IDB)

Lecture 11: Multisets and Aggregation

Multisets

Multiset sets where the same element can occur multiples times (SQL uses multisets)
- **Multiplicity** is the number of occurences of an element.
- **Bags** another name for multisets.

Operation to remove multiples (ϵ)

As described.

Basic SQL

```
Q := SELECT [DISTINCT] a FROM t WHERE c
    | Q1 UNION [ALL] Q2
    | Q1 INTERSECT [ALL] Q2
    | Q1 EXCEPT [ALL] Q2
```

`SELECT a FROM t WHERE c` – keeps duplicates, `[DISTINCT]` removes them
`UNION, INTERSECT & EXCEPT` – remove duplicates, `[ALL]` keeps them

SQL to RA on bags

SQL	RA on bags
`SELECT` α ...	$\pi_\alpha(.)$
`SELECT DISTINCT` α ...	$\epsilon(\pi_\alpha(.))$
Q_1 `UNION ALL` Q_2	$Q_1 \cup Q_2$
Q_1 `INTERSECT ALL` Q_2	$Q_1 \cap Q_2$
Q_1 `EXCEPT ALL` Q_2	$Q_1 - Q_2$
Q_1 `UNION` Q_2	$\epsilon(Q_1 \cup Q_2)$
Q_1 `INTERSECT` Q_2	$\epsilon(Q_1 \cap Q_2)$
Q_1 `EXCEPT` Q_2	$\epsilon(Q_1) - Q_2$

- duplicates are good because they give you a true distribution of the data

Aggregate Functions in SQL

`COUNT` number of elements in a column
`AVG` average value of all elements in column
`SUM` Adds up all elements in a column

`MIN / MAX` min/max values of elements in a column
`COUNT (*)` counts all rows in table
`COUNT (DISTINCT *)` is ILLEGAL! use, `SELECT COUNT(DISTINCT T.*)`