

Multisets

Dr Paolo Guagliardo



THE UNIVERSITY *of* EDINBURGH
informatics

v26.0

Changelog

v26.0 Initial version

Duplicates

R	$\pi_A(R)$	SELECT A FROM R
<u>A B</u>	<u>A</u>	<u>A</u>
a1 b1	a1	a1
a2 b2	a2	a2
a1 b2		a1

- ▶ We considered relational algebra on **sets**
- ▶ SQL uses **bags**: sets with duplicates

Multisets (a.k.a. bags)

Sets where the **same element** can occur **multiple times**

The **number of occurrences** of an element is called its **multiplicity**

Notation

$a \in_k B$: a occurs k times in bag B

$a \in B$: a occurs in B with multiplicity ≥ 1

$a \notin B$: a does not occur in B (that is, $a \in_0 B$)

Relational algebra on bags

Relations are **bags of tuples**

Projection

Keeps duplicates

$$\pi_A \left(\begin{array}{cc} \mathbf{A} & \mathbf{B} \\ \hline 2 & 3 \\ 1 & 1 \\ 2 & 2 \end{array} \right) = \begin{array}{c} \mathbf{A} \\ \hline 2 \\ 1 \\ 2 \end{array}$$

Relational algebra on bags

Cartesian product

Concatenates tuples as many times as they occur

$$\begin{array}{cc} \mathbf{A} & \mathbf{B} \\ \hline 1 & 1 \end{array} \times \begin{array}{c} \mathbf{C} \\ \hline 2 \\ 2 \end{array} = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ \hline 1 & 1 & 2 \\ 1 & 1 & 2 \end{array}$$

Relational algebra on bags

Selection

Takes all occurrences of tuples satisfying the condition:

$$\text{If } \bar{a} \in_k R, \quad \text{then } \begin{cases} \bar{a} \in_k \sigma_\theta(R) & \text{if } \bar{a} \text{ satisfies } \theta \\ \bar{a} \notin \sigma_\theta(R) & \text{otherwise} \end{cases}$$

Example

$$\sigma_{A>1} \left(\begin{array}{c|cc} \mathbf{A} & \mathbf{B} \\ \hline 2 & 3 \\ 1 & 2 \\ 2 & 3 \end{array} \right) = \begin{array}{c|cc} \mathbf{A} & \mathbf{B} \\ \hline 2 & 3 \\ 2 & 3 \end{array}$$

Relational algebra on bags

Duplicate elimination ε

New operation that removes duplicates:

$$\text{If } \bar{a} \in R, \quad \text{then } \bar{a} \in_1 \varepsilon(R)$$

Example

$$\varepsilon \left(\begin{array}{c|cc} \mathbf{A} & \mathbf{B} \\ \hline 2 & 3 \\ 1 & 2 \\ 2 & 3 \end{array} \right) = \begin{array}{c|cc} \mathbf{A} & \mathbf{B} \\ \hline 2 & 3 \\ 1 & 2 \end{array}$$

Relational algebra on bags

Union

Adds multiplicities:

If $\bar{a} \in_k R$ and $\bar{a} \in_n S$, then $\bar{a} \in_{k+n} R \cup S$

Example

A	B		A	B		A	B
1	2		1	2		1	2
1	2	∪	1	3	=	1	2
1	3		1	4		1	3
						1	2
						1	3
						1	4

Relational algebra on bags

Intersection

Takes the **minimum** multiplicity:

If $\bar{a} \in_k R$ and $\bar{a} \in_n S$, then $\bar{a} \in_{\min\{k,n\}} R \cap S$

Example

A	B		A	B		A	B
1	2		1	2		1	2
1	2	∩	1	3	=	1	3
1	3		1	4			

Relational algebra on bags

Difference

Subtracts multiplicities up to zero:

$$\text{If } \bar{a} \in_k R \text{ and } \bar{a} \in_n S, \text{ then } \begin{cases} \bar{a} \in_{k-n} R - S & \text{if } k > n \\ \bar{a} \notin R - S & \text{otherwise} \end{cases}$$

Example

A	B		A	B		A	B
1	2	−	1	2	=	1	2
1	2		1	3			
1	3		1	3			
			1	4			

RA on sets vs. RA on bags

Equivalences of RA on sets **do not** necessarily hold on **bags**

Example

On bags $\sigma_{\theta_1 \vee \theta_2}(R) \not\equiv \sigma_{\theta_1}(R) \cup \sigma_{\theta_2}(R)$

R	A	$\sigma_{A>1 \vee A<3}(R)$	A	$\sigma_{A>1}(R) \cup \sigma_{A<3}(R)$	A
	2		2		2
					2

$$\varepsilon(\sigma_{\theta_1 \vee \theta_2}(R)) \equiv \varepsilon(\sigma_{\theta_1}(R) \cup \sigma_{\theta_2}(R)) \text{ holds}$$

Basic SQL queries revisited

$$\begin{aligned} Q := & \text{SELECT } [\text{DISTINCT}] \alpha \text{ FROM } \tau \text{ WHERE } \theta \\ & | Q_1 \text{ UNION } [\text{ALL}] Q_2 \\ & | Q_1 \text{ INTERSECT } [\text{ALL}] Q_2 \\ & | Q_1 \text{ EXCEPT } [\text{ALL}] Q_2 \end{aligned}$$

SQL and RA on bags

SQL

SELECT $\alpha \dots$
SELECT DISTINCT $\alpha \dots$

Q_1 **UNION ALL** Q_2
 Q_1 **INTERSECT ALL** Q_2
 Q_1 **EXCEPT ALL** Q_2

Q_1 **UNION** Q_2
 Q_1 **INTERSECT** Q_2
 Q_1 **EXCEPT** Q_2

RA on bags

$\pi_\alpha(\cdot)$
 $\varepsilon(\pi_\alpha(\cdot))$

$Q_1 \cup Q_2$
 $Q_1 \cap Q_2$
 $Q_1 - Q_2$

$\varepsilon(Q_1 \cup Q_2)$
 $\varepsilon(Q_1 \cap Q_2)$
 $\varepsilon(Q_1) - Q_2$