Database Management Systems

Mathematical Preliminaries

Malay Bhattacharyya

Assistant Professor

Machine Intelligence Unit
Indian Statistical Institute, Kolkata

January, 2019

Preliminaries

Relation: Given the sets $X_1, X_2, \ldots, X_n \subseteq \mathbb{R}$ (the real plane), a relation $\mathcal{R}$ can be defined on $X_1, X_2, \ldots, X_n$ as $\mathcal{R} = \{(x_1, x_2, \ldots, x_n) : (x_1, x_2, \ldots, x_n) \in X_1 \times X_2 \times \cdots \times X_n\}$.

If the sets denote different attributes in a database then a table represents nothing but a relation (subset of the Cartesian product of attributes) between the attributes.

Based on this, we can assume:
A **relation** is a table
The **attributes** are the headers of the table
A **tuple** is a row.

Preliminaries

Example of a relation:

Table: MATH_OLYMPIC

Year	Gold	Silver
2008	0	0
2009	0	3
2010	0	2
2011	1	1
2012	2	3
2013	0	2
2014	0	1
2015	0	1
2016	0	1
2017	0	0
2018	0	3

Preliminaries

Example of another relation:

Table: MATH_OLYMPIC_GOLDEN

Year	Gold	Silver
2011	1	1
2012	2	3

Preliminaries of relational algebra

Query language: A language for manipulation and retrieval of data from a database.

* Query languages can be – procedural (user provides requirements along with instructions) and non-procedural (user provides requirements only).

The relational algebra is a procedural query language

The relational algebra works on relations

Note: Relational algebra is *closed* because every operation in relational algebra returns a relation.

Preliminaries of relational algebra

Relational algebra is not "Turing complete". This is inevitably favourable because it menifests that relational algebra is subject to algorithmic analysis (to be precise for query optimization).

Union

Notation: $R_1 \cup R_2$, where R_1, R_2 are relational algebra expressions.

Description: Returns tuples that appear in either or both of the two relations, thereby producing a relation with at most $\mathcal{T}(R_1) + \mathcal{T}(R_2)$ tuples.

Note: Union operation is valid iff the attributes of R_1 and R_2 are the same, i.e. $\mathcal{A}(R_1) = \mathcal{A}(R_2)$.

Union

Example: MATH_OLYMPIC $\cup$ MATH_OLYMPIC_GOLDEN

Year	Gold	Silver
2008	0	0
2009	0	3
2010	0	2
2011	1	1
2012	2	3
2013	0	2
2014	0	1
2015	0	1
2016	0	1
2017	0	0
2018	0	3

Intersection

Notation: $R_1 \cap R_2$, where R_1, R_2 are relational algebra expressions.

Description: Returns tuples that appear in both the relations, thereby producing a relation with at most $\min(\mathcal{T}(R_1), \mathcal{T}(R_2))$ tuples.

Note: Intersection operation is valid iff the attributes of R_1 and R_2 are the same, i.e. $\mathcal{A}(R_1) = \mathcal{A}(R_2)$.

Intersection

Example: MATH_OLYMPIC $\cap$ MATH_OLYMPIC_GOLDEN

Year	Gold	Silver
2011	1	1
2012	2	3

Difference

Notation: $R_1 - R_2$, where R_1, R_2 are relational algebra expressions.

Description: Returns the tuples that appear in one relation (first one) but not in the other (second one), thereby producing a relation with at most $\mathcal{T}(R_1)$ tuples.

Note: Difference operation is valid iff the attributes of R_1 and R_2 are the same, i.e. $\mathcal{A}(R_1) = \mathcal{A}(R_2)$.

Difference

Example: MATH_OLYMPIC − MATH_OLYMPIC_GOLDEN

Year	Gold	Silver
2008	0	0
2009	0	3
2010	0	2
2013	0	2
2014	0	1
2015	0	1
2016	0	1
2017	0	0
2018	0	3

Cartesian product

Notation: $R_1 \times R_2$, where R_1, R_2 are relational algebra expressions.

Description: Returns the Cartesian product of two relations, thereby producing a relation with attributes $\mathcal{A}(R_1) \cup \mathcal{A}(R_2)$ and $\mathcal{T}(R_1) * \mathcal{T}(R_2)$ number of tuples.

Note: No validity constraint.

Cartesian product

Example: MATH_OLYMPIC $\times$ MATH_OLYMPIC_GOLDEN

M.Year	M.Gold	M.Silver	M_G.Year	M_G.Gold	M_G.Silver
2008	0	0	2011	1	1
2008	0	0	2012	2	3
2009	0	3	2011	1	1
2009	0	3	2012	2	3
2010	0	2	2011	1	1
2010	0	2	2012	2	3
2011	1	1	2011	1	1
2011	1	1	2012	2	3
2012	2	3	2011	1	1
2013	2	3	2012	2	3
...	...	...	...	...	...
2018	0	3	2011	1	1
2018	0	3	2012	2	3

Selection

Notation: $\sigma_P(R)$, where P is a predicate on the attributes of the relation R.

Description: Returns the tuples that satisfy a given predicate (extracts a subset of tuples).

Selection

Example: $\sigma_{\text{Gold}\neq 0}$(MATH_OLYMPIC)

Year	Gold	Silver
2011	1	1
2012	2	3

Example: $\sigma_{\text{Gold}\neq 0 \wedge \text{Silver}>1}$(MATH_OLYMPIC)

Year	Gold	Silver
2012	2	3

Projection

Notation: $\pi_S(R)$, where S is a subset of the attributes in the relation R.

Description: Returns all tuples with the given attributes only (extracts a subset of attributes).

Projection

Example: $\pi_{\text{Year,Silver}}$(MATH_OLYMPIC)

Year	Silver
2008	0
2009	3
2010	2
2011	1
2012	3
2013	2
2014	1
2015	1
2016	1
2017	0
2018	3

Example: $\pi_{\text{Year,Silver}}(\sigma_{\text{Gold}>1}$(MATH_OLYMPIC))

Year	Silver
2012	3

Rename

Notation: $\rho_N(R)$, where N is the new name for the result of R.

Description: Renames a relation in relational algebra.

Rename

Example: $\rho_{IMO}(\text{MATH\_OLYMPIC})$

Table: IMO

Year	Gold	Silver
2008	0	0
2009	0	3
2010	0	2
2011	1	1
2012	2	3
2013	0	2
2014	0	1
2015	0	1
2016	0	1
2017	0	0
2018	0	3

Natural join

Notation: $R_1 \bowtie R_2$, where R_1, R_2 are relations obtained from relational algebra operations.

Description: Cartesian product of two relations followed by the removal of duplicate attributes.

Note: If we consider the pair of relations R_1 and R_2, then the natural join between them $(R_1 \bowtie R_2)$ is a relation on schema $\mathcal{A}(R_1) \cup \mathcal{A}(R_2))$ such that

$$R_1 \bowtie R_2 = \pi_{\mathcal{A}(R_1) \cup \mathcal{A}(R_2))}(\sigma_{\mathcal{A}_1(R_1)=\mathcal{A}_1(R_2) \wedge \ldots \wedge \mathcal{A}_n(R_1)=\mathcal{A}_n(R_2)} R_1 \times R_2).$$

Natural join

Example: $\pi_{\text{Year}}(\text{IMO} \bowtie \text{MATH\_OLYMPIC\_GOLD})$

Year
2011
2012

Theta join

Notation: $R_1 \bowtie_\theta R_2$, where R_1, R_2 are relations obtained from relational algebra operations.

Description: Cartesian product of two relations followed by selection operation. We can write

$$R_1 \bowtie_\theta R_2 = \sigma_\theta(R_1 \times R_2).$$

Note: The result of theta join is defined only if the attributes of the relations are disjoint.

EQUI join

Notation: $R_1 \bowtie_{=} R_2$, where R_1, R_2 are relations obtained from relational algebra operations.

Description: Cartesian product of two relations followed by selection operation with respect to equity. EQUI join is a special case of theta join where $\theta =$ "$=$".

Division

Notation: $R_1 \div R_2$, where R_1, R_2 are relations obtained from relational algebra operations.

Description: Satisfies universal specification.

Note: Division operation is valid iff the attributes of R_2 is a proper subset of R_1, i.e. $\mathcal{A}(R_2) \subseteq \mathcal{A}(R_1)$. A tuple is said to be in $R_1 \div R_2$ iff the tuple is in $\pi_{\mathcal{A}(R_1)-\mathcal{A}(R_2)}(R_1)$ and its Cartesian product with any arbitrary tuple in R_2 produces a tuple that belongs to R_1. Interestingly, we can represent the division operation as follows

$$\begin{aligned} R_1 \div R_1 &= \pi_{\mathcal{A}(R_1)-\mathcal{A}(R_2)}(R_1) - \\ &\quad \pi_{\mathcal{A}(R_1)-\mathcal{A}(R_2)}((\pi_{\mathcal{A}(R_1)-\mathcal{A}(R_2)}(R_1) \times R_2) - R_1). \end{aligned}$$

Division

Example: Let us consider the following pair of relations.

Table: CODE

Roll	Coding	Feature
1	Python	Programming
2	C	Programming
2	R	Programming
3	Python	Programming
3	Python	Visualization
4	C++	Programming
5	R	Visualization

Table: SKILL

Feature
Programming
Visualization

CODE $\div$ SKILL

Roll	Coding
3	Python

Assignment

Notation: $var \leftarrow R$, where var is a variable and R is a relation obtained from relational algebra operations

Description: Assigns a relational algebra expression to a relational variable

Example: Gold $\leftarrow E$

Outer join – Basics

Outer join has been extended from the natural join operation for avoiding information loss. Let us consider the following pair of relations.

Table: FAC

Name	Unit	Centre
Malay	MIU	Kolkata
Mandar	CVPRU	Kolkata
Ansuman	ACMU	Kolkata
Sandip	ACMU	Kolkata

Table: RES

Name	Area	Level
Malay	CB	Junior
Mandar	IR	Senior
Sasthi	WSN	Senior
Sandip	DM	Senior

Outer join – Motivation

Example: FAC $\bowtie$ RES

Name	**Unit**	**Centre**	**Area**	**Level**
Malay	MIU	Kolkata	CB	Junior
Mandar	CVPRU	Kolkata	IR	Senior
Sandip	ACMU	Kolkata	DM	Senior

The information about *Ansuman* and *Sasthi* are lost.

Outer join – Left outer join

Notation: $R_1 ⟕ R_2$, where R_1, R_2 are relations obtained from relational algebra operations

Description: Makes a natural join but adds extra tuples to the result padded with NULL values to deal with missing information in the first relation

Outer join – Left outer join

Example: FAC $⟕$ RES

Name	Unit	Centre	Area	Level
Malay	MIU	Kolkata	CB	Junior
Mandar	CVPRU	Kolkata	IR	Senior
Sandip	ACMU	Kolkata	DM	Senior
Ansuman	ACMU	Kolkata	NULL	NULL

Outer join – Right outer join

Notation: $R_1 ⟖ R_2$, where R_1, R_2 are relations obtained from relational algebra operations

Description: Makes a natural join but adds extra tuples to the result padded with NULL values to deal with missing information in the second relation

Outer join – Right outer join

Example: FAC ⟖ RES

Name	Unit	Centre	Area	Level
Malay	MIU	Kolkata	CB	Junior
Mandar	CVPRU	Kolkata	IR	Senior
Sandip	ACMU	Kolkata	DM	Senior
Sasthi	NULL	NULL	WSN	Senior

Outer join – Full outer join

Notation: $R_1 ⟗ R_2$, where R_1, R_2 are relations obtained from relational algebra operations

Description: Makes a natural join but adds extra tuples to the result padded with NULL values to deal with missing information in both the relations

Outer join – Full outer join

Example: FAC $⟗$ RES

Name	Unit	Centre	Area	Level
Malay	MIU	Kolkata	CB	Junior
Mandar	CVPRU	Kolkata	IR	Senior
Sandip	ACMU	Kolkata	DM	Senior
Ansuman	ACMU	Kolkata	NULL	NULL
Sasthi	NULL	NULL	WSN	Senior

Completeness

A complete set comprises a subset of relational algebra operations that can express any other relational algebra operations.
E.g., the set $\{\sigma, \pi, \cup, -, \times\}$ is complete.