RELATIONAL DATABASES

GROUP ACTIVITY

THEORETICAL CONTENTS

N-ary relations

A table that has two columns can be understood as a binary relation of A on B, where A is a set composed of the elements in the first column and B is the set composed of the elements in the second column.

The elements of a binary relation, which is defined by a table with two columns, can be represented as ordered pairs (a, b) of $A \times B$. The first coordinate corresponds to the first column and the second coordinate corresponds to the second column. The two coordinates of an ordered pair are in the same row.

$A \times B$		*A*	*B*
$\vdots$		$\vdots$	$\vdots$
(a, b)	$\leftrightarrow$	a	b
$\vdots$		$\vdots$	$\vdots$

When we have a table with n columns, we can similarly define a relation. In this case, there are as many sets (i.e., $A_1, A_2, \ldots, A_n$) as columns in the table (i.e., n columns). The first set is composed of the elements in the first column, the second set is composed of the elements in the second column, …, and the n-set is composed of the elements in the n-column. A n-ary relation is a set of $A_1 \times A_2 \times \ldots \times A_n$ which is composed by n-tuples. The order of the element in a tuple is the same than the order of the elements in a row of the table.

$A_1 \times A_2 \times \ldots \times A_n$		*A1*	*A2*		*An*
$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$
$(a_1, a_2, \ldots, a_n)$	$\leftrightarrow$	a1	a2	…	an
$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$

Table example 1 FILMS R_1

Identification Number	*Title*	*Year*	*Director*
22013	The Shawshank Redemption	1994	Frank Darabont
93833	The Godfather	1972	Francis Ford Coppola
57197	The Godfather Part II	1974	Francis Ford Coppola
89765	The Dark Knight	2008	Christopher Nolan
26785	12 Angry Men	1957	Sidney Lumet
61258	Schindler's List	1993	Steven Spielberg

24006	The Lord of the Rings: The Return of the King	2003	Peter Jackson
25369	Pulp Fiction	1994	Quentin Tarantino

Table example 2 GENRES R_2

Identification Number of Films	*Genre*
22013	Drama
23456	Crime
35678	Comedy
57197	Crime
59087	Thriller
61258	Drama

Relational Databases

A ***database*** is just an ordered collection of data which is managed by a computer. ***Database management systems*** are programs that allows the users to access of the information stored in the databases.

The ***relational model*** of E.F. Codd (1970) is based on the n-ary relations to define the structure (tables) of the relational databases. For example, a relational database could be composed of the previous relations R_1 (see Table example 1) and R_2 (see Table example 2).

In a relational database, the columns of the n-ary relations are called ***attributes*** (or fields). For example, the attributes of the relation R_1 are *Identification Number, Title, Year,* and *Director*. In contrast, the attributes of the relation R_2 are *Identification Number of Films,* and *Genre*.

Moreover, the domain of an attribute is the set to which all its elements (the elements of the column) belong. For example, the domain of the attribute *Year* is an integer between 1895 (the first film was released) and today.

Each tuple in a relation is uniquely defined by a **key**, so that there are no two exactly equal tuples. The key can be composed of one or more attributes of the relation. For example, in the relation R_1, we can have two films with same *Title* (e.g., Cinderella), with the same *Year* (e.g., 1994), or with the same *Director* (e.g., Francis Ford Coppola). However, the *Identification number* would be different for each film. Therefore, we can use the attribute *Identification number* as the key to the relation R_1. We could also select a combination of attributes as key. For example, the *Title* and *Year* of a film or the *Title* and *Director*.

A database management system is based on ***queries***. A query is a request to obtain some information that is stored in the database. For example, if we want to know what films were released in 1994, we will write a query to obtain this information from the database.

Operations with relational databases

The following relational operations are used to design and write the queries for relational databases.

1. Selection $\sigma_{\text{condition}}(\text{Name of the Relation})$

The selection operator allows to select some specific tuples (rows) of a relation.

The tuples are selected depending on conditions.

Example

For the relation R_1 (FILMS) of the Table example 1, the following expression selects the tuples whose attribute *Year* is equal to 1994:

$$\sigma_{\text{Year}=1994}(\text{FILMS})$$

This expression will return the following 4-tuples: (22013, The Shawshank Redemption, 1994, Frank Darabont) and (25369, Pulp Fiction, 1994, Quentin Tarantino). In other words, it will return the relation given by this table:

Identification Number	*Title*	*Year*	*Director*
22013	The Shawshank Redemption	1994	Frank Darabont
25369	Pulp Fiction	1994	Quentin Tarantino

Therefore, it selects the tuples of the relation *FILMS,* which satisfies a specific condition: Year=1994.

2. Projection $\pi_{(\text{attributes separated by ;})}(\text{Name of the Relation})$

The projection operator allows to select the elements of specific columns and it discards the duplicated values.

To specify the column or columns to be projected, we must use the name of the correspondent attribute or attributes.

Example

For the relation R_1 (FILMS) of the Table example 1, the following expression projects the second and the fourth columns (i.e., Title and Director):

$$\pi_{(\text{Title;Year})}(\text{FILMS})$$

This expression will return the set of 2-tuples:

$\pi_{(\mathrm{Title;Year})}(\mathrm{FILMS})$ = { (The Shawshank Redemption, 1994), (The Godfather, 1972), (The Godfather Part II, 1974), (The Dark Knight, 2008), (12 Angry Men, 1957), (Schindler's List, 1993), (The Lord of the Rings: The Return of the King, 2003), (Pulp Fiction, 1994) }

In other words, it will return the relation given by this table:

Title	*Year*
The Shawshank Redemption	1994
The Godfather	1972
The Godfather Part II	1974
The Dark Knight	2008
12 Angry Men	1957
Schindler's List	1993
The Lord of the Rings: The Return of the King	2003
Pulp Fiction	1994

Note that there are no duplicated tuples. In case multiple tuples are equal, we discard all but one.

3. Join $R_1 \bowtie_{(\mathrm{condition})} R_2$, where$R_1$ and R_2 are the names of the relations (tables).

Given two relations or tables R_1 and R_2, the join operator compares a specific condition to them. This condition is called join condition and specifies the relation between an attribute in the first table and an attribute in the second table.

If the condition is satisfied, the n-tuples are combined to form a new relation, which is the solution of the query. In this new relation, the tuples of the first relation R_1 will correspond to the first columns, and then, the following columns will contain the tuples of the second relation R_2. The duplicated tuples must be discarded.

Example

Given the relations R_1 y R_2 for *FILMS* and *GENRES* (see Table example 1 and Table example 2), we can join these relations using the following condition:

Identification Number = Identification Number of Films

To do this, we write the following expression:

$$R_1 \bowtie_{(\mathrm{Identification\ Number\ =\ Identification\ Number\ of\ Films})} R_2$$

or

$$\mathrm{FILMS} \bowtie_{(\mathrm{Identification\ Number\ =\ Identification\ Number\ of\ Films})} \mathrm{GENRES}$$

The result will be the following tuples:

(22013, The Shawshank Redemption, 1994, Frank Darabont, Drama), (57197, The Godfather Part II, 1974, Francis Ford Coppola, Crime), and (61258, Schindler's List, 1993, Steven Spielberg, Drama)

In other way, it will return the following table which defines a new relation:

Identification Number	*Title*	*Year*	*Director*	*Genre*
22013	The Shawshank Redemption	1994	Frank Darabont	Drama
57197	The Godfather Part II	1974	Francis Ford Coppola	Crime
61258	Schindler's List	1993	Steven Spielberg	Drama

Comments

✓ The Selection operator and the Projection operator are always applied on one relation, so they are called unary operators. In contrast, the Join operator is applied on two relations, so it is called binary operator.

✓ As you can see, the result of a query is a new table, that means, a new relation. To use the results of queries, it is a good practice to assign a name to the result of our queries.

✓ In relational databases, most of the queries require several simultaneously operations to provide an answer.

Examples

1. The previous queries could be written as follows:

$$C_1 := \sigma_{\text{Year}=1994}(\text{FILMS})$$

$$C_2 := \pi_{(\text{Title;Year})}(\text{FILMS})$$

$$C_3 := \text{FILMS} \bowtie_{(\text{Identification Number} = \text{Identification Number of Films})} \text{GENRES}$$

2. To do the following query: *Determine the titles of all the drama films,* we have to do three operations. First, we join the relations which contains the titles of the films (R_1 called FILMS) and the genres (R_2 called GENRES). Then, we select the tuples whose genre is 'Drama'. Finally, we do a projection of the attribute *Title* to return the titles of the films.

$$C_4 := \pi_{(\text{Title})}\left(\sigma_{Genre=Drama}(\text{FILMS} \bowtie_{(\text{Identification Number= Identification Number of Films})} \text{GENRES})\right)$$

The result will be the following 1-tuples: (The Shawshank Redemption), (Schindler's List)

In other words, we will obtain relation C_4, given by the following table:

Title
The Shawshank Redemption
Schindler's List

EXERCISES

Given the following relations or tables, solve the following exercises.

Table 1: MUSICIANS (R_1)

Name	*Number of studio albums*	*Nationality*
Dua Lipa	2	British
Maluma	5	Colombian
Ana Mena	1	Spanish
David Guetta	7	French
The Weeknd	4	Canadian
Nathy Peluso	1	Argentinean

Table 4: STAGES (R_4)

Stage name	*Capacity (number of attendees)*	*City*
WiZink Center	17400	Madrid
Palau Sant Jordi	17900	Barcelona
Auditorio Starlite	3000	Marbella
Soldeu Avet slope platform	5000	Soldeu (Andorra)
Razzmatazz	300	Barcelona
Guíxols Arena	400	Sant Feliu de Guixols
Sala Apolo	1300	Barcelona

Table 2: CONCERTS (R_2)

Title	*Musician name*
Future Nostalgia EU TOUR 2022	Dua Lipa
Maluma – 11:11 World Tour	Maluma
David Guetta	David Guetta
The After Hours Tour	The Weeknd
Nathy Peluso	Nathy Peluso

Table 3: PROGRAMMING (R_3)

Stage	*Date*	*Concert title*
WiZink Center	03-06-2022 21:00	Future Nostalgia EU TOUR 2022
Palau Sant Jordi	01-06-2022 21:00	Future Nostalgia EU TOUR 2022
WiZink Center	05-04-2022 21:00	Maluma – 11:11 World Tour
Auditorio Starlite	22-06-2022 22:00	Maluma
Soldeu Avet slope platform	31-07-2021 21:00	David Guetta
Palau Sant Jordi	28-10-2022 21:30	The After Hours Tour
Auditorio Starlite	12-08-2021 22:00	Nathy Peluso

Exercise 1. Express the sets of the attributes in the previous tables indicating their domains. Then, express the relations in these tables as set of n-tuples indicating their cartesian products.

Exercise 2. Write the necessary operation to answer the following queries. Specify the new relation that is the solution of the operations using a table or a set of n-tuples.

a) Determine the name of the musicians that have released only one studio album.

b) Determine all the number of studio albums.

c) Determine all the concerts in Madrid.

d) Determine the names of the musicians and the titles of the concerts programmed for June 2022, that is, between 01-06-2022 00:00 and 30-06-2022 23:59.

Exercise 3. Write the necessary operation to answer the following queries. Specify the new relation that is the solution of the operations using a table or a set of n-tuples.

a) Determine all the dates when there is a concert.

b) Determine the title and the date of the concerts that will take place in the "WiZink Center".

c) Determine the names of the musicians who will perform in the stage "Palau Sant Jordi".

d) Determine the names of the musicians who will perform in a stage of Barcelona.

Exercise 4. Write the necessary operation to answer the following queries. Specify the new relation that is the solution of the operations using a table or a set of n-tuples.

a) Determine all the cities where there is some stage.

b) Determine the name of the stage whose capacity is greater than 1000 and is in Barcelona.

c) Determine the title and dates of all the concerts that are programmed in Barcelona.

d) Determine the names of the musicians who will perform in a stage not located in Barcelona.

Exercise 5. Write the necessary operation to answer the following queries. Specify the new relation that is the solution of the operations using a table or a set of n-tuples.

a) Determine the nationalities of all the musicians.

b) Determine the names and nationalities of all the musicians who have released more than 3 studio albums.

c) Determine the nationality of the musicians whose concerts have the name of the artists as title.

d) Determine the nationality of the musicians that will perform in Barcelona.

Exercise 6. Considering a relational database as a n-ary relation expressed with a table:

a) Describe the operator ***union*** over the relation database.

b) Provide an example about how this operator works answering the following query: *Determine the title of the concerts that will take place in Madrid or in Barcelona.* Write the necessary operations to answer this query.

Exercise 7. Considering a relational database as a n-ary relation expressed with a table:

a) Describe the operator ***intersection*** over the relation database.

b) Provide an example about how this operator works answering the following query: *Determine the title of the concerts that will take place in Madrid and in Barcelona.* Write the necessary operations to answer this query.

Exercise 8. Considering a relational database as a n-ary relation expressed with a table:

a) Describe the operator ***difference*** over the relation database.

b) Provide an example about how this operator works answering the following query: *Determine the title of the concerts that will not take place in Madrid.* Write the necessary operations to answer this query.

Note that to solve the proposed query you could also define a 'complement operator' over the relation database.

Exercise 9. Considering a relational database as a n-ary relation expressed with a table:

a) Describes the operator ***symmetric difference*** over the relation database.

b) Provide an example about how this operator works answering the following query: *Determine the title of the concerts that will take place in Madrid or in Barcelona, but they will not take place in both stages.* Write the necessary operations to answer this query.

References

Notes of the course.

R. Johnsonbaugh, "Discrete mathematics". Prentice Hall, 1997