Database Systems

Relational Algebra

H. Turgut Uyar Şule Öğüdücü

2002-2015

License

Topics

Closure

- closure: input and output of all operations are relations
- output of one operation can be the input to another
- operations can be nested

Example Database

MOVIE

MOVIE#	TITLE	YEAR	SCORE	VOTES	DIRECTOR#
6	The Usual Suspects	1995	8.7	35027	639
70	Being John Malkovich	1999	8.3	13809	1485
107	Batman & Robin	1997	3.5	10577	105
110	Sleepy Hollow	1999	7.5	10514	148
112	Three Kings	1999	7.7	10319	1070
151	Gattaca	1997	7.4	8388	2020
213	Blade	1998	6.7	6885	2861
228	Ed Wood	1994	7.8	6587	148
251	End of Days	1999	5.5	6095	103
281	Dangerous Liaisons	1988	7.7	5651	292
373	Fear and Loathing in Las Vegas	1998	6.5	4658	59
432	Stigmata	1999	6.1	4141	2557
433	eXistenZ	1999	6.9	4130	97
573	Dead Man	1995	7.4	3333	175
1468	Europa	1991	7.6	1042	615
1512	Suspiria	1977	7.1	1004	2259
1539	Cry-Baby	1990	5.9	972	364

Example Database

PERSON

PERSON#	NAME
9	Arnold Schwarzenegger
26	Johnny Depp
59	Terry Gilliam
97	David Cronenberg
103	Peter Hyams
105	Joel Schumacher
138	George Clooney
148	Tim Burton
175	Jim Jarmusch
187	Christina Ricci
243	Uma Thurman
282	Cameron Diaz
292	Stephen Frears
302	Benicio Del Toro
308	Gabriel Byrne
350	Jennifer Jason Leigh

364	John Waters
406	Patricia Arquette
503	John Malkovich
615	Lars von Trier
639	Bryan Singer
745	Udo Kier
793	Jude Law
1070	David O. Russell
1485	Spike Jonze
1641	Iggy Pop
2020	Andrew Niccol
2259	Dario Argento
2557	Rupert Wainwright
2861	Stephen Norrington
3578	Traci Lords

Example Database

CASTING

MOVIE#	ACTOR#	ORD
6	308	2
6	302	3
70	282	2
70	503	14
107	9	1
107	138	2
107	243	4
110	26	1
110	187	2
112	138	1
112	1485	4
151	243	2
151	793	3
213	745	6
213	3578	8
228	26	1
228	406	4
251	9	1
251	308	2

251	745	10
281	243	7
281	503	2
373	26	1
373	187	6
373	282	8
373	302	2
432	308	2
432	406	1
433	350	1
433	793	2
573	26	1
573	308	12
573	1641	6
1468	745	3
1512	745	9
1539	26	1
1539	1641	5
1539	3578	7

Selection

- selection: selecting tuples that satisfy a condition

  `relation WHERE condition`

- output header = input header

Selection Examples - 1

- movies with more than 10000 votes (S1)

```
MOVIE WHERE (VOTES > 10000)
```

S1

MOVIE#	TITLE	YEAR	SCORE	VOTES	DIRECTOR#
6	The Usual Suspects	1995	8.7	35027	639
70	Being John Malkovich	1999	8.3	13809	1485
107	Batman & Robin	1997	3.5	10577	105
110	Sleepy Hollow	1999	7.5	10514	148
112	Three Kings	1999	7.7	10319	1070

Selection Examples - 2

- movies older than 1992, with scores higher than 7.5 (S2)

```
MOVIE WHERE ((YEAR < YEAR(1992))
        AND (SCORE > SCORE(7.5)))
```

S2

MOVIE#	TITLE	YEAR	SCORE	VOTES	DIRECTOR#
281	Dangerous Liaisons	1988	7.7	5651	292
1468	Europa	1991	7.6	1042	615

Projection

- projection: selecting a set of attributes

```
relation { attribute_name [, ...] }
```

- output header = attribute list

Projection Examples - 1

- titles of all movies (P1)

```
MOVIE { TITLE }
```

P1

TITLE
The Usual Suspects
Being John Malkovich
Batman & Robin
Sleepy Hollow
Three Kings
Gattaca
Blade
Ed Wood
End of Days

Dangerous Liaisons
Fear and Loathing in Las Vegas
Stigmata
eXistenZ
Dead Man
Europa
Suspiria
Cry-Baby

Projection Examples - 2

- titles and years of all movies (P2)

  `MOVIE { TITLE, YEAR }`

P2

TITLE	YEAR
Batman & Robin	1997
Being John Malkovich	1999
Blade	1998
Cry-Baby	1990
Dangerous Liaisons	1988
Dead Man	1995
Ed Wood	1994
End of Days	1999
Europa	1991
Fear and Loathing in Las Vegas	1998
Gattaca	1997
Sleepy Hollow	1999
Stigmata	1999
Suspiria	1977
Three Kings	1999
The Usual Suspects	1995
eXistenZ	1999

Projection Examples - 3

- years of all movies (P3)

  `MOVIE { YEAR }`

P3

YEAR
1995
1999
1997
1998
1994
1988
1991
1977
1990

Projection Examples - 4

- titles of movies with votes more than 5000 and scores higher than 7.0 (P4)

1. movies with votes more than 5000 and scores higher than 7.0 (P4A)
2. titles in P4A (P4)

Projection Examples - 4

- movies with votes more than 5000 and scores higher than 7.0 (P4A)

```
MOVIE WHERE ((VOTES > 5000)
        AND (SCORE > SCORE(7.0)))
```

P4A

MOVIE#	TITLE	YEAR	SCORE	VOTES	DIRECTOR#
6	The Usual Suspects	1995	8.7	35027	639
70	Being John Malkovich	1999	8.3	13809	1485
110	Sleepy Hollow	1999	7.5	10514	148
112	Three Kings	1999	7.7	10319	1070
151	Gattaca	1997	7.4	8388	2020
228	Ed Wood	1994	7.8	6587	148
281	Dangerous Liaisons	1988	7.7	5651	292

Projection Examples - 4

- titles in P4A (P4)

```
P4A { TITLE }
```

P4

TITLE
Being John Malkovich
Dangerous Liaisons
Ed Wood
Gattaca
Sleepy Hollow
Three Kings
The Usual Suspects

Projection Examples - 4

- titles of movies with votes more than 5000 and scores higher than 7.0 (P4)

```
( MOVIE
    WHERE ((VOTES > 5000)
       AND (SCORE > SCORE(7.0))) )
  { TITLE }
```


Join

- join: matching tuples of two relations over common values of one or more attributes

```
relation1 JOIN relation2
```

- natural join: over common values of all the attributes with the same name

Join

- from the Cartesian product of the two relations, selecting the tuples which have the same values for the given attributes
- matching attributes are not repeated in the output
- output header = relation1 header $\cup$ relation2 header

Join Examples - 1

- titles and director names of all movies (J1)

1. all movies and their directors (J1A)
2. titles and director names in J1A (J1)

Join Examples - 1

- all movies and their directors (J1A)

```
MOVIE JOIN
  (PERSON RENAME { PERSON# AS DIRECTOR# })
```

J1A

MOVIE#	TITLE	...	DIRECTOR#	NAME
6	The Usual Suspects	...	639	Bryan Singer
70	Being John Malkovich	...	1485	Spike Jonze
107	Batman & Robin	...	105	Joel Schumacher
...	...	...	...	...
1468	Europa	...	615	Lars von Trier
1512	Suspiria	...	2259	Dario Argento
1539	Cry-Baby	...	364	John Waters

Join Examples - 1

- titles and director names in J1A (J1)

```
J1A { TITLE, NAME }
```

J1

TITLE	NAME
Batman & Robin	Joel Schumacher
Being John Malkovich	Spike Jonze
Blade	Stephen Norrington
...	...
Three Kings	Spike Jonze
The Usual Suspects	Bryan Singer
eXistenZ	David Cronenberg

Join Examples - 2

- titles, actor names and actor orders of all movies (J2)

1. all movies and their casting data (J2A)
2. all data in J2A matched with persons (J2B)
3. titles, actor names and actor orders in J2B (J2)

Join Examples - 2

- all movies and their casting data (J2A)

```
MOVIE JOIN CASTING
```

J2A

MOVIE#	TITLE	...	ACTOR#	ORD
6	The Usual Suspects	...	302	3
6	The Usual Suspects	...	308	2
70	Being John Malkovich	...	282	2
70	Being John Malkovich	...	503	14
...	...	...	...	...
1539	Cry-Baby	...	26	1
1539	Cry-Baby	...	1641	5
1539	Cry-Baby	...	3578	7

Join Examples - 2

- all data in J2A matched with persons (J2B)

```
J2A JOIN
  (PERSON RENAME { PERSON# AS ACTOR# })
```

J2B

MOVIE#	TITLE	...	ACTOR#	ORD	NAME
6	The Usual Suspects	...	302	3	Benicio Del Toro
6	The Usual Suspects	...	308	2	Gabriel Byrne
70	Being John Malkovich	...	282	2	Cameron Diaz
70	Being John Malkovich	...	503	14	John Malkovich
...	...	...	...	...	...
1539	Cry-Baby	...	26	1	Johnny Depp
1539	Cry-Baby	...	1641	5	Iggy Pop
1539	Cry-Baby	...	3578	7	Traci Lords

Join Examples - 2

- titles, actor names and actor orders in J2B (J2)

```
J2B { TITLE, NAME, ORD }
```

J2

TITLE	NAME	ORD
The Usual Suspects	Benicio Del Toro	3
The Usual Suspects	Gabriel Byrne	2
Being John Malkovich	Cameron Diaz	2
Being John Malkovich	John Malkovich	14
...	...	...
Cry-Baby	Johnny Depp	1
Cry-Baby	Iggy Pop	5
Cry-Baby	Traci Lords	7

Join Examples - 3

- names of the actors in Johnny Depp's movies (J3)

1. ids of Johnny Depp's movies (J3A)
2. ids of the actors in the movies in J3A (J3B)
3. names of the actors in J3B (J3)

Join Examples - 3

- ids of Johnny Depp's movies (J3A)

```
(((PERSON RENAME { PERSON# AS ACTOR# })
    JOIN CASTING)
  WHERE (NAME = "Johnny Depp")) { MOVIE# }
```

J3A

MOVIE#
110
228
373
573
1539

Join Examples - 3

- ids of the actors in the movies in J3A (J3B)

```
(J3A JOIN CASTING) { ACTOR# }
```

J3B

ACTOR#
26
187
282
302
308
406
1641
3578

Join Examples - 3

- names of the actors in J3B (J3)

```
((J3B RENAME { ACTOR# AS PERSON# })
    JOIN PERSON) { NAME }
```

J3

NAME
Johnny Depp
Christina Ricci
Cameron Diaz
Benicio Del Toro
Gabriel Byrne
Patricia Arquette
Iggy Pop
Traci Lords

Division

- division: from the tuples of the first relation, selecting the ones that match all the tuples of the second relation in a mediating relation

```
relation1 DIVIDEBY relation2
  PER (relation3)
```

Division Example

- titles of movies with Johnny Depp and Christina Ricci (V1)

1. ids of Johnny Depp and Christina Ricci (V1A)
2. ids of movies with all the actors in V1A (V1B)
3. titles of the movies in V1B (V1)

Division Example

- ids of Johnny Depp and Christina Ricci (V1A)

```
(PERSON
    WHERE ((NAME = "Johnny Depp")
        OR (NAME = "Christina Ricci")))
  { PERSON# }
```

V1A

PERSON#
26
187

Division Example

- ids of movies with all the actors in V1A (V1B)

```
(MOVIE { MOVIE# })
    DIVIDEBY
(V1A RENAME { PERSON# AS ACTOR# })
    PER (CASTING { MOVIE#, ACTOR# })
```

V1B

MOVIE#
110
373

Division Example

- titles of movies in V1B (V1)

```
(V1B JOIN MOVIE) { TITLE }
```

V1

TITLE
Fear and Loathing in Las Vegas
Sleepy Hollow

Division Example

- product - division relationship:
  `V1B JOIN V1A` $\subseteq$ `CASTING { MOVIE#, ACTOR# }`

MOVIE#	ACTOR#
110	26
110	187
373	26
373	187

Intersection

- intersection: selecting the tuples found in both relations

  `relation1 INTERSECT relation2`

- output header = relation1 header = relation2 header

Intersection Example

- names of all directors who also acted (`I1`)

1. ids of all directors who also acted (`I1A`)
2. names of all the persons in `I1A` (`I1`)

Intersection Example

- ids of all directors who also acted (`I1A`)

```
(MOVIE { DIRECTOR# }
      RENAME { DIRECTOR# AS PERSON# })
  INTERSECT
(CASTING { ACTOR# }
      RENAME { ACTOR# AS PERSON# })
```

I1A

PERSON#
1485

Intersection Example

- names of all persons in `I1A` (`I1`)

  `(I1A JOIN PERSON) { NAME }`

I1

NAME
Spike Jonze

Union

- union: selecting the tuples found in at least one of two relations

  `relation1 UNION relation2`

- output header = relation1 header = relation2 header

Union Example

- names of all directors and actors of all movies newer than 1997 (`U1`)

1. ids and director ids of all movies newer than 1997 (`U1A`)
2. ids of all actors in the movies in `U1A` (`U1B`)
3. ids of directors and actors in at least one of `U1A` and `U1B` (`U1C`)
4. names of all the persons in `U1C` (`U1`)

Union Example

- ids and director ids of all movies newer than 1997 (`U1A`)

```
(MOVIE WHERE (YEAR > YEAR(1997)))
  { MOVIE#, DIRECTOR# }
```

U1A

MOVIE#	DIRECTOR#
70	1485
110	148
112	1070
213	2861
251	103
373	59
432	2557
433	97

Union Example

- ids of all actors in the movies in U1A (U1B)

```
(U1A JOIN CASTING) { ACTOR# }
```

U1B

ACTOR#
9
26
138
187
282
302
308
350
406
503
745
793
1485
3578

Union Example

- ids of directors and actors in at least one of U1A and U1B (U1C)

```
(U1A { DIRECTOR# }
      RENAME { DIRECTOR# AS PERSON# })
UNION (U1B RENAME { ACTOR# AS PERSON# })
```

U1C

PERSON#
9
26
59
97
103
138
148
187
282
302
308
350
406
503
745
793
1070
1485
2557
2861
3578

Union Example

- names of all persons in U1C (U1)

```
(U1C JOIN PERSON) { NAME }
```

U1

NAME
Arnold Schwarzenegger
Benicio Del Toro
Cameron Diaz
Christina Ricci
David Cronenberg
David O. Russell
Gabriel Byrne
George Clooney
Jennifer Jason Leigh
John Malkovich
Johnny Depp
Jude Law
Patricia Arquette
Peter Hyams
Rupert Wainwright
Spike Jonze
Stephen Norrington
Terry Gilliam
Tim Burton
Traci Lords
Udo Kier

Difference

- difference: selecting the tuples which are found in the first but not in the second relation

```
relation1 MINUS relation2
```

- output header = relation1 header = relation2 header

Difference Example

- names of all actors who haven't played in Johnny Depp's movies (`D1`)

1. ids of all actors who played in Johnny Depp's movies (`J3B`)
2. names of all actors who are not in `J3B` (`D1`)

Difference Example

- names of all the actors who are not in `J3B` (`D1`)

```
(((CASTING { ACTOR# } MINUS J3B)
      RENAME { ACTOR# AS PERSON# })
  JOIN PERSON) {NAME}
```

D1

NAME
Arnold Schwarzenegger
George Clooney
Jennifer Jason Leigh
John Malkovich
Jude Law
Spike Jonze
Udo Kier
Uma Thurman

References

Required Reading: Date

- Chapter 7: Relational Algebra
  7.1. Introduction
  7.2. Closure Revisited
  7.4. The Original Algebra: Semantics

Projection

- selecting columns from a table:

```
SELECT [ ALL | DISTINCT ] column_name [, ...]
  FROM table_name
```

- duplicate rows are allowed
  - `ALL`: preserve duplicate rows (default)
  - `DISTINCT`: take only one of duplicate rows
- $*$: all columns

Query Examples

- all data of all movies

```
SELECT * FROM MOVIE
```

- titles and years of all movies

```
SELECT TITLE, YR FROM MOVIE
```

- years when movies were released

```
SELECT DISTINCT YR FROM MOVIE
```


Ordering Results

- ordering the rows in the result

```
SELECT [ ALL | DISTINCT ] column_name [, ...]
  FROM table_name
  [ ORDER BY { column_name [ ASC | DESC ] }
      [, ...] ]
```

- ASC: in ascending order (default)
- DESC: in descending order

Query Examples

- years when movies were released, in ascending order

```
SELECT DISTINCT YR FROM MOVIE
  ORDER BY YR
```

- years when movies were released, in descending order

```
SELECT DISTINCT YR FROM MOVIE
  ORDER BY YR DESC
```


Query Examples

- all data of movies ordered first by descending years then by ascending titles

```
SELECT * FROM MOVIE
  ORDER BY YR DESC, TITLE ASC
```

Column Expressions

- evaluating expressions over columns

```
SELECT [ ALL | DISTINCT ]
  { expression [ AS column_name ] } [, ...]
  FROM table_name
  [ ORDER BY { column_name [ ASC | DESC ] }
      [, ...] ]
```

- new columns can be named: AS
- name or index of column can be used for ordering

Query Examples

- titles and total scores of all movies

```
SELECT TITLE, SCORE * VOTES
  FROM MOVIE
```


Query Examples

- titles and total scores of all movies,
  in descending order of total scores

```
SELECT TITLE, SCORE * VOTES AS POINTS
  FROM MOVIE
  ORDER BY POINTS DESC
```

```
SELECT TITLE, SCORE * VOTES
  FROM MOVIE
  ORDER BY 2 DESC
```


Selection

- selecting rows from a table

```
SELECT [ ALL | DISTINCT ]
  { expression [ AS column_name ] } [, ...]
  FROM table_name
  [ WHERE condition ]
  [ ORDER BY { column_name [ ASC | DESC ] }
      [, ...] ]
```

Query Examples

- year of "Citizen Kane"

```
SELECT YR FROM MOVIE
  WHERE (TITLE = 'Citizen Kane')
```

- titles of movies with scores less than 3 and votes more than 10

```
SELECT TITLE FROM MOVIE
  WHERE ((SCORE < 3) AND (VOTES > 10))
```


Condition Expressions

- check if null:

```
column_name IS { NULL | NOT NULL }
```

- set membership:

```
column_name IN (value_set)
```

- string comparison:

```
column_name LIKE pattern
```

- %: matches any group of symbols, including empty string

Query Examples

- titles of movies with unknown year

```
SELECT TITLE FROM MOVIE
  WHERE (YR IS NULL)
```

- titles and years of movies in the years 1967, 1954 and 1988

```
SELECT TITLE, YR FROM MOVIE
  WHERE (YR IN (1967, 1954, 1988))
```


Query Examples

- titles and scores of "Police Academy" movies

```
SELECT TITLE, SCORE FROM MOVIE
  WHERE (TITLE LIKE 'Police Academy%')
```

Grouping

- grouping selected rows

```
SELECT [ ALL | DISTINCT ]
  { expression [ AS column_name ] } [, ...]
  FROM table_name
  [ WHERE condition ]
  [ GROUP BY column_name [, ...] ]
  [ HAVING condition ]
  [ ORDER BY { column_name [ ASC | DESC ] }
      [, ...] ]
```

- result contains one row per group
- groups can be filtered: HAVING

Processing Order

- select the rows that satisfy the WHERE condition
- group the selected rows using the columns specified in the GROUP BY clause
  - if no group, the entire result is one group
- select the groups that satisfy the HAVING condition
- calculate the expressions in the column list
- order the result on the columns specified in the ORDER BY clause

Group Values

- one value for each group
- column used for grouping
- result of an aggregate function on a column
- aggregate functions: COUNT SUM AVG MAX MIN
- column name as parameter
- null values are ignored

Query Examples

- for every year, the number of movies with score greater than 8.5 in that year

```
SELECT YR, COUNT(ID) FROM MOVIE
  WHERE (SCORE > 8.5)
  GROUP BY YR
```

- score of the favorite movie of every year, in ascending order of years

```
SELECT YR, MAX(SCORE) FROM MOVIE
  GROUP BY YR
  ORDER BY YR
```

Query Examples

- total number of votes

```
SELECT SUM(VOTES) FROM MOVIE
```


Query Examples

- averages of movie scores
  in the years where there are at least 25 movies
  for which more than 40 people have voted,
  in ascending order of years

```
SELECT YR, AVG(SCORE)
   FROM MOVIE
   WHERE (VOTES > 40)
   GROUP BY YR
   HAVING (COUNT(ID) >= 25)
   ORDER BY YR
```


Join

- joining can be achieved using `WHERE` conditions
- list tables to join `FROM` clause
- use dotted notation for columns with identical names
- processing order (conceptual):
  - take Cartesian product of the tables
  - select rows that satisfy the `WHERE` condition
  - ...

Query Examples

- name of the director of “Star Wars”

```
SELECT NAME
  FROM MOVIE, PERSON
  WHERE ((DIRECTORID = PERSON.ID)
     AND (TITLE = 'Star Wars'))
```

Query Examples

- names of the actors in "Alien"

```
SELECT NAME
  FROM MOVIE, PERSON, CASTING
  WHERE ((TITLE = 'Alien')
     AND (MOVIEID = MOVIE.ID)
     AND (ACTORID = PERSON.ID))
```


Query Examples

- titles of Harrison Ford's movies

```
SELECT TITLE
  FROM MOVIE, PERSON, CASTING
  WHERE ((NAME = 'Harrison Ford')
     AND (MOVIEID = MOVIE.ID)
     AND (ACTORID = PERSON.ID))
```


Query Examples

- titles and names of the lead actors of the movies in 1962

```
SELECT TITLE, NAME
  FROM MOVIE, PERSON, CASTING
  WHERE ((YR = 1962)
     AND (MOVIEID = MOVIE.ID)
     AND (ACTORID = PERSON.ID)
     AND (ORD = 1))
```


Query Examples

- number of movies John Travolta acted in every year

```
SELECT YR, COUNT(MOVIEID)
   FROM MOVIE, PERSON, CASTING
   WHERE ((NAME = 'John Travolta')
      AND (MOVIEID = MOVIE.ID)
      AND (ACTORID = PERSON.ID))
   GROUP BY YR
```

Query Examples

- titles and number of actors of the movies in 1978, in descending order of actor counts

```
SELECT TITLE, COUNT(ACTORID)
   FROM MOVIE, CASTING
   WHERE ((YR = 1978)
      AND (MOVIE.ID = CASTING.MOVIEID))
   GROUP BY MOVIEID, TITLE
   ORDER BY 2 DESC
```


Table Expressions

- joins can be expressed as table expressions:

```
SELECT ...
  FROM table_expression [ AS table_name ]
  WHERE selection_condition
  ...
```

- product
- using conditions
- over columns with the same name
- natural join
- outer join

Join Expressions

- product

```
SELECT ...
  FROM table1_name CROSS JOIN table2_name
  ...
```

- join using conditions

```
SELECT ...
  FROM table1_name JOIN table2_name
         ON condition
  ...
```


Query Examples

- name of the director of “Star Wars”

```
SELECT NAME
  FROM MOVIE, PERSON
  WHERE ((DIRECTORID = PERSON.ID)
     AND (TITLE = 'Star Wars'))

SELECT NAME
  FROM MOVIE JOIN PERSON
          ON (DIRECTORID = PERSON.ID)
  WHERE (TITLE = 'Star Wars')
```

Join Expressions

- over columns with the same name

```
SELECT ...
  FROM table1_name JOIN table2_name
         USING (column_name [, ...])
  ...
```

- repeated columns are taken once
- natural join

```
SELECT ...
  FROM table1_name NATURAL JOIN table2_name
  ...
```


Outer Join

- inner join: unmatched rows are excluded
- outer join: unmatched rows are included
- columns from the other table are null

```
SELECT ...
  FROM table1_name [ LEFT | RIGHT | FULL ]
         [ OUTER ] JOIN table2_name
  ...
```


Outer Join Examples

- left outer join

T1

NUM	NAME
1	a
2	b
3	c

T2

NUM	VALUE
1	xxx
3	yyy
5	zzz

SELECT * FROM T1 LEFT JOIN T2

NUM	NAME	NUM	VALUE
1	a	1	xxx
2	b		
3	c	3	yyy

Outer Join Examples

- right outer join

T1

NUM	NAME
1	a
2	b
3	c

T2

NUM	VALUE
1	xxx
3	yyy
5	zzz

SELECT * FROM T1 RIGHT JOIN T2

NUM	NAME	NUM	VALUE
1	a	1	xxx
3	c	3	yyy
		5	zzz

Outer Join Examples

- full outer join

T1

NUM	NAME
1	a
2	b
3	c

T2

NUM	VALUE
1	xxx
3	yyy
5	zzz

SELECT * FROM T1 FULL JOIN T2

NUM	NAME	NUM	VALUE
1	a	1	xxx
2	b		
3	c	3	yyy
		5	zzz

Query Examples

- titles of movies with no known actors

```
SELECT TITLE
  FROM MOVIE LEFT JOIN CASTING
         ON (MOVIEID = MOVIE.ID)
  WHERE (ACTORID IS NULL)
```


Self Join

- how to join columns in the same table?
- give a new name to the table in the expression

Query Examples

- titles of all movies with the same number of votes

```
SELECT M1.TITLE, M2.TITLE
  FROM MOVIE AS M1, MOVIE AS M2
  WHERE (M1.VOTES = M2.VOTES)
    AND (M1.ID < M2.ID)
```

Subqueries

- using subquery results in condition expressions

```
SELECT ...
  WHERE expression operator
      [ ALL | ANY ] (subquery)
  ...
```

- result of subquery must be compatible
- ALL: for all values from subquery
- ANY: for at least one value from subquery

Query Examples

- titles and scores of all movies with scores
  higher than that of "Star Wars", in descending order of scores

```
SELECT TITLE, SCORE FROM MOVIE
   WHERE (  SCORE >
      ( SELECT SCORE FROM MOVIE
          WHERE (TITLE = 'Star Wars') )
) ORDER BY SCORE DESC
```


Query Examples

- titles of movies with scores less than
  the scores of all "Police Academy" movies

```
SELECT TITLE FROM MOVIE
   WHERE (  SCORE < ALL
      ( SELECT SCORE FROM MOVIE
          WHERE (TITLE LIKE 'Police Academy%') )
)
```


Query Examples

- titles of movies with less votes
  than any movie made before 1930

```
SELECT TITLE FROM MOVIE
   WHERE ((YR >= 1930) AND (  VOTES < ANY (
      ( SELECT VOTES FROM MOVIE
          WHERE (YR < 1930) )
))
```

Query Examples

- names of actors who played with Johnny Depp

```
SELECT NAME FROM PERSON, CASTING
   WHERE ((ACTORID = PERSON.ID)
      AND (MOVIEID IN (
           SELECT MOVIEID
              FROM PERSON, CASTING
              WHERE ((ACTORID = PERSON.ID)
                 AND (NAME = 'Johnny Depp'))
) ))
```


Query Examples

- names of actors with at least 10 lead roles

```
SELECT NAME FROM PERSON
   WHERE (ID IN (
         SELECT ACTORID FROM CASTING
            WHERE (ORD = 1)
            GROUP BY ACTORID
            HAVING (COUNT(MOVIEID) >= 10)
) )
```


Set Operations

- an operation on two subquery results
- intersection: `INTERSECT`
- union: `UNION`
- difference: `EXCEPT`
- duplicate rows are not allowed in the result

Query Examples

- number of people who both directed and acted

```
SELECT COUNT(*) FROM (
   ( SELECT DISTINCT DIRECTORID FROM MOVIE )
   INTERSECT
   ( SELECT DISTINCT ACTORID FROM CASTING )
) AS DIRECTOR_ACTOR
```

Query Examples

- number of people who worked in movies before 1930

```
SELECT COUNT(*) FROM (
   ( SELECT DISTINCT DIRECTORID FROM MOVIE
       WHERE (YR < 1930) )
   UNION
   ( SELECT DISTINCT ACTORID FROM CASTING
       WHERE (MOVIEID IN
                ( SELECT ID FROM MOVIE
                    WHERE (YR < 1930) )) )
) AS OLD_MOVIE_PERSON_IDS
```


Query Examples

- number of directors who have not acted

```
SELECT COUNT(*) FROM (
   ( SELECT DISTINCT DIRECTORID FROM MOVIE )
   EXCEPT
   ( SELECT DISTINCT ACTORID FROM CASTING )
) AS DIRECTOR_ONLY
```


References

Required Reading: Date

- Chapter 8: Relational Calculus
  - 8.6. SQL Facilities
- Appendix B: SQL Expressions
- Chapter 19: Missing Information

Supplementary Reference

- A Gentle Introduction to SQL:
  `http://sqlzoo.net/`