**Advanced SQL Queries on IMDB**

**JOINS**

**List the number of titles that were produced before 1990 and had at least one**

**person in the cast that was named Brad Pitt.**

SELECT COUNT(titles.start\_year)

FROM titles

LEFT JOIN cast\_info ON titles.title\_id = cast\_info.title\_id

LEFT JOIN persons ON cast\_info.person\_id = persons.person\_id

WHERE titles.start\_year < 1990 AND persons.full\_name = 'Brad Pitt'

**List all names of people that have produced at least 5 different titles**

SELECT persons.full\_name

FROM persons

INNER JOIN cast\_info ON cast\_info.person\_id = persons.person\_id AND cast\_info.job\_category = 'producer'

INNER JOIN titles ON titles.title\_id = cast\_info.title\_id

GROUP BY persons.full\_name

HAVING COUNT(titles.title\_id) > 4

**List all job categories and (possibly) associated character names of the person**

**named ‘Keanu Reeves’, e.g (‘actor’, ‘Neo’), ....**

SELECT DISTINCT cast\_info.job\_category, title\_person\_character.character\_name

FROM cast\_info, title\_person\_character

INNER JOIN persons ON persons.person\_id = title\_person\_character.person\_id

WHERE persons.person\_id = cast\_info.person\_id AND persons.full\_name = 'Keanu Reeves'

**For each (of the 28) title genres stored in the database, list the number of titles**

**named (primary title) ‘The Interview’, e.g. (‘Documentary’, 7), .... (Hint: First**

**construct the table containing all the different ‘The Interview’ titles)**

SELECT titles\_genres.genre, COUNT(titles.primary\_title)

FROM titles\_genres

INNER JOIN titles ON titles\_genres.title\_id = titles.title\_id

WHERE titles.primary\_title = 'The Interview'

GROUP BY titles\_genres.genre

**SETS**

**List all distinct (primary) title names including a character named ‘Chewbacca’**

**except for those produced before 2015, e.g. (‘Star Wars: Detours’), ....**

SELECT titles.primary\_title

FROM titles

INNER JOIN (

SELECT title\_id

FROM titles

WHERE start\_year > 2015

INTERSECT

SELECT title\_person\_character.title\_id

FROM title\_person\_character

WHERE character\_name = 'Chewbacca'

) AS matching\_titles

ON titles.title\_id = matching\_titles.title\_id;

**List the names of all people that have been both directors and actors/actresses**

**anytime in their career, e.g. (‘Demi Moore’), ...**

SELECT DISTINCT persons.full\_name

FROM persons

JOIN cast\_info ON cast\_info.person\_id = persons.person\_id

WHERE cast\_info.job\_category = 'director'

INTERSECT

SELECT DISTINCT persons.full\_name

FROM persons

JOIN cast\_info ON cast\_info.person\_id = persons.person\_id

WHERE cast\_info.job\_category IN ('actor','actress')

**SUBQUERIES**

**List the number of people for whom at least another person with the same full**

**name exists, but with a different id. Use the EXISTS clause.**

SELECT COUNT(p.person\_id)

FROM persons p

WHERE EXISTS(

SELECT persons.full\_name

FROM persons

WHERE p.full\_name = persons.full\_name AND p.person\_id != persons.person\_id

)

**List the primary title(s) with the second largest number of actresses (not**

**actors) in its cast. Hint: Use WITH clause to construct a table counting the**

**number of actresses for each title.**

WITH num\_actresses AS (

SELECT titles.title\_id, COUNT(cast\_info.job\_category) as actress\_count

FROM titles, cast\_info

WHERE titles.title\_id = cast\_info.title\_id AND cast\_info.job\_category = 'actress'

GROUP BY titles.title\_id

)

SELECT MAX(titles.primary\_title)

FROM titles

JOIN num\_actresses ON num\_actresses.title\_id = titles.title\_id

WHERE num\_actresses.actress\_count < (

SELECT MAX(num\_actresses.actress\_count)

FROM num\_actresses

)

**Solve query 5 with a subquery. Make sure that the new query returns the same**

**results as query 5.**

I already used a subquery

**Write a with-clause that returns a table (called distinct\_types) storing all**

**distinct title types. Afterwards, write a query that lists the title genres that are**

**associated with titles of all such types (types stored in distinct\_types).**

WITH distinct\_types AS (

SELECT DISTINCT titles.title\_type

FROM titles

)

SELECT DISTINCT titles\_genres.genre

FROM titles\_genres,titles

WHERE titles\_genres.genre IN (

SELECT distinct\_types.title\_type

FROM distinct\_types

)

GROUP BY titles\_genres.genre

HAVING COUNT(DISTINCT titles.title\_type) = (SELECT COUNT(\*) FROM distinct\_types)