

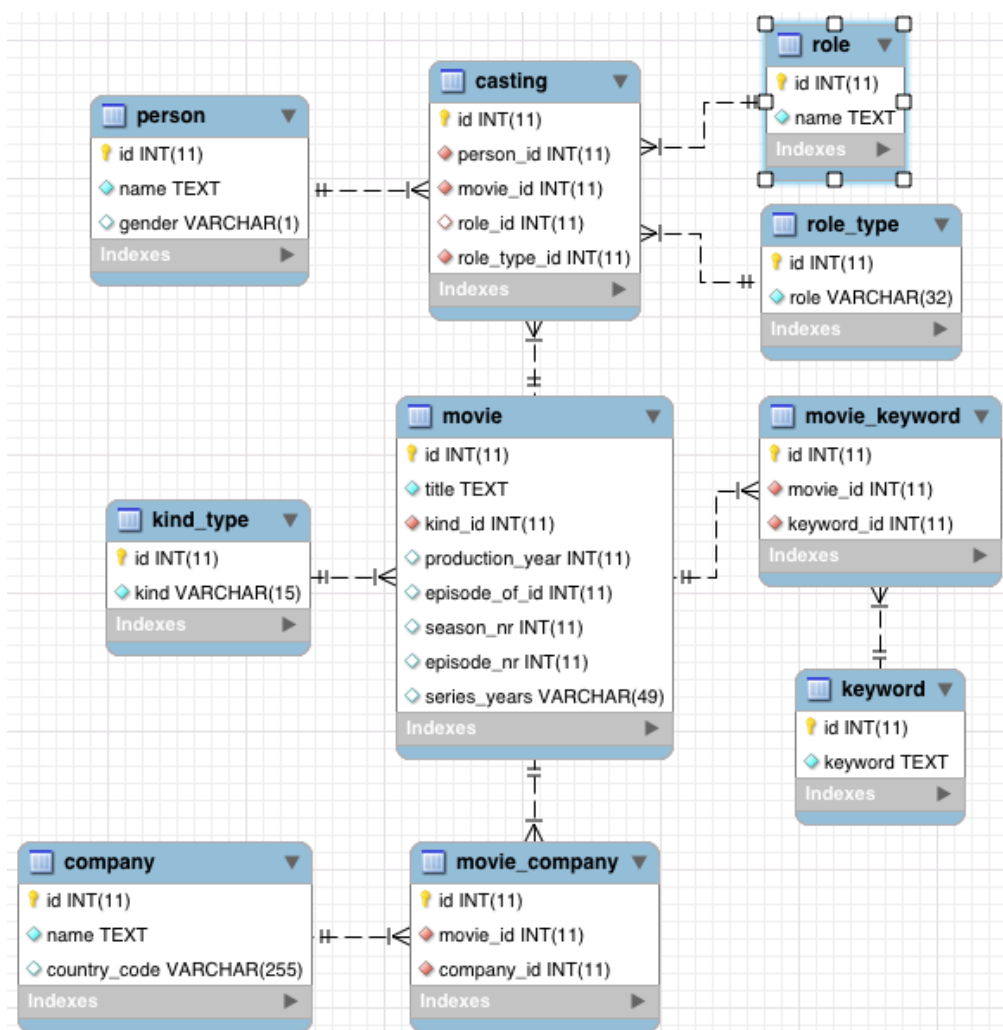
RAWDATA Assignment 1 – Querying IMDB with SQL

The main purpose of this assignment is to practice querying using SQL. The target is a huge dump of “real world” data from IMDB covering more than 3.5 million titles stored in a relational database **imdb_movie** on the server **islb.ruc.dk**. Use the connection you made for the **movie** database last week or consult the note “MySQL Installation” if you need to set up a new connection. When your connection is open execute the command “use imdb_movie;” to make this schema default thereby avoiding inclusion of schema-name prefix, when referring to tables.

The available **imdb_movie** database is only an excerpt of what can be downloaded from IMDB and the tables are simplified and renamed. Compare the relational schema and the diagram below to get an overview of the database. The relational schema includes the following 10 tables.

| Table name | No of rows |
|---------------|------------|
| casting | 49462688 |
| role | 3810797 |
| company | 302225 |
| keyword | 186008 |
| kind_type | 7 |
| movie_company | 3897767 |
| movie_keyword | 6103575 |
| person | 5428172 |
| role_type | 12 |
| movie | 3570524 |

The data model is the following.



Responsive Applications, Web services and Databases

The database schema is shown as a so-called crow's foot diagram. All connections are one-to-many. For instance, the connection between person and casting clarifies that one casting connects one person, while a person can be connected to zero, one or more (many) castings. Notice that the arrows all represent a referring attribute in the origin table (named "..._id"). All tables have a unique attribute "id". Thus, one way (among several) to express a join involving e.g. person and casting can be by applying a constraint `casting.person_id = person.id`.

What to do

Write SQL queries to answer the questions below. Notice that the results are given, so you will "simply" have to find an SQL-expression, that produces this result.

How and when to hand in

Please collect all your query-expressions in a single script (in a single query-window). When you have tested these one by one, do the following:

- In the MySQL Workbench **Query** menu, choose **Execute (All or Selection) to Text**.
- This will process all your queries and produce answers as text output. Copy the output to a text-file and
- hand in this file (upload the file on Moodle "Assignment 1" no later than February 13).

Important

Group hand in is accepted, but DO REMEMBER to write ALL NAMES of participants in your group in the top of the text file. ONLY one version of your result needs to be uploaded.

Questions

a) Find the titles and production year of all movies with a title starting with 'Pirates of the Caribbean'.

| title | production_year |
|--|-----------------|
| Pirates of the Caribbean 6 | NULL |
| Pirates of the Caribbean: At World's End | 2007 |
| Pirates of the Caribbean: Dead Man's Chest | 2006 |
| Pirates of the Caribbean: Dead Men Tell No Tales | 2017 |
| Pirates of the Caribbean: On Stranger Tides | 2011 |
| Pirates of the Caribbean: Tales of the Code: Wedlocked | 2011 |
| Pirates of the Caribbean: The Curse of the Black Pearl | 2003 |
| Pirates of the Caribbean: The Young Sparrow | 2016 |

8 rows in set (0.01 sec)

b) How many movies were produced in 2004?

| COUNT(*) |
|----------|
| 14143 |

1 row in set (0.01 sec)

c) Find the title of all video games with Mads Mikkelsen.

| title |
|-------------------|
| Quantum of Solace |

1 row in set (0.01 sec)

d) Find the different roles that Kevin Bacon has had.

| role |
|------|
|------|

Henrik Bulskov & Troels Andreasen

Responsive Applications, Web services and Databases

| |
|-----------------|
| actor |
| director |
| producer |
| writer |
| cinematographer |
| editor |

```
+-----+
6 rows in set (0.01 sec)
```

e) Find for each role the number of persons assigned to the roles order descending by the number.

| role | number |
|---------------------|----------|
| actor | 16622527 |
| actress | 9906721 |
| miscellaneous crew | 6170099 |
| producer | 5936394 |
| writer | 3799109 |
| director | 2353225 |
| editor | 1599479 |
| cinematographer | 1234113 |
| composer | 1055730 |
| production designer | 421641 |
| costume designer | 363650 |

```
+-----+
11 rows in set (0.01 sec)
```

f) Find the title of titles directed by Ridley Scott from 2004, 2006, 2008 or 2010.

| |
|--------------|
| title |
| A Good Year |
| Body of Lies |
| Robin Hood |

```
+-----+
3 rows in set (0.01 sec)
```

g) What is the highest number of persons casted for one title?

| |
|--------|
| MAX(c) |
| 1742 |

```
+-----+
1 row in set (0.01 sec)
```

h) How many actors has Kevin Bacon acted together with?

| |
|--------------------|
| my_count_of_actors |
| 5721 |

```
+-----+
1 row in set (0.01 sec)
```

i) Find the title of all movies that are assigned to the keywords 'sister' and 'elephant'

| |
|---------------------|
| title |
| The Diamond of Doom |
| Ang Panday 2 |
| O'Shaughnessy's Boy |
| The Circus Show-Up |
| The Hugga Bunch |

Responsive Applications, Web services and Databases

+-----+

5 rows in set (0.01 sec)

j) Find the title and production year of movies involving the company "Paramount" in Sweden produced after 2004 ordered by title.

| title | production_year |
|----------------|-----------------|
| Choke | 2008 |
| Dreamgirls | 2006 |
| Hotelliggaren | 2005 |
| Hustle & Flow | 2005 |
| Madagascar | 2005 |
| Over the Hedge | 2006 |
| Stardust | 2007 |

+-----+
 7 rows in set (0.01 sec)

k) Find the title of movies that have casted to a role named "The Singing Kid"

| title |
|--------|
| Broken |

+-----+
 1 row in set (0.05 sec)

l) Who have played a role named "Bilbo" in a movie where the word "Smaug" appear in the title.

| name | title |
|-----------------|---|
| Baxter, Daniel | How the Desolation of Smaug Should Have Ended |
| Freeman, Martin | The Hobbit: The Desolation of Smaug |