

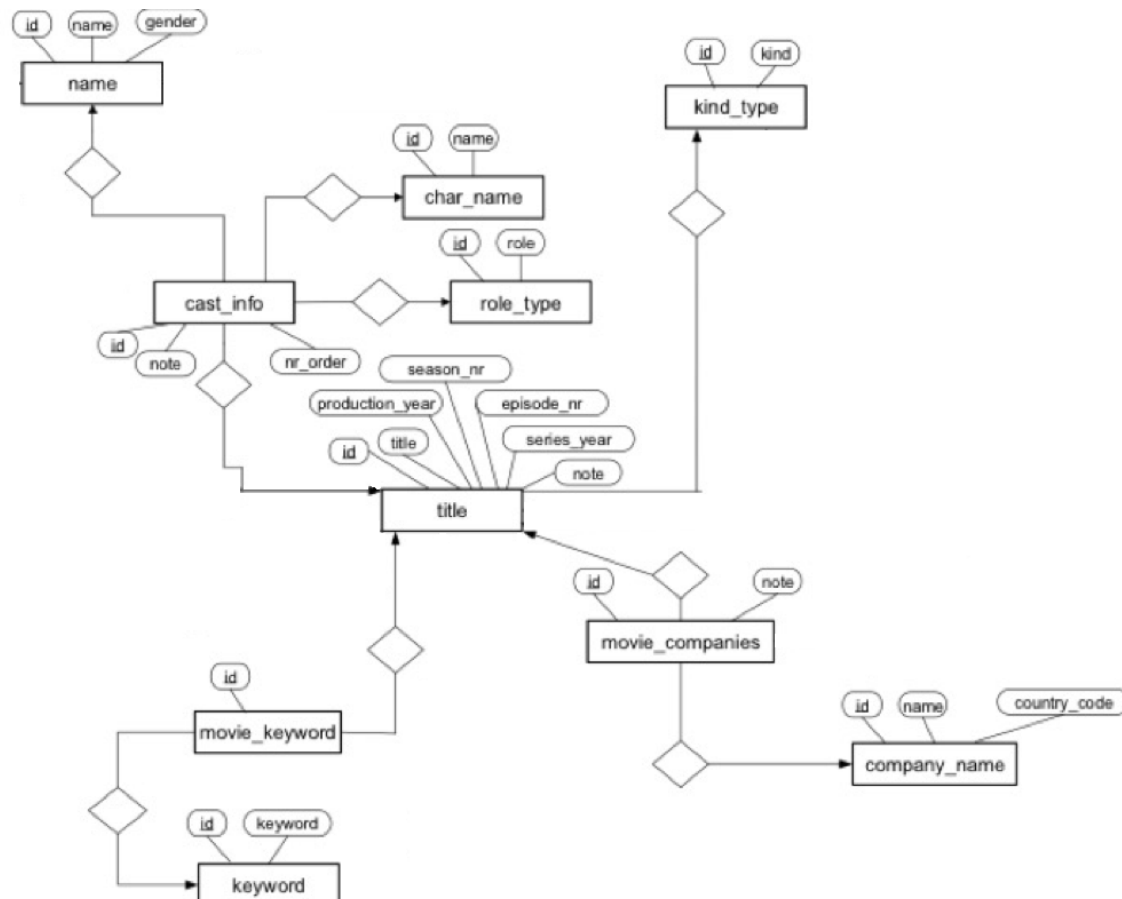
RAWDATA Assignment 1 – Querying IMDB with SQL

The main purpose of this assignment is to practice querying using SQL. The target is a huge dump of “real world” data from IMDB (from early 2016) covering more than 3.5 million titles stored in a relational database **imdb2016** 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 imdb2016;” to avoid inclusion of the schema-name prefix, when referring to tables.

The full IMDB data model is quite complex but you will not need to go into every detail. Compare the relational schema and the diagram below to get an overview of the database. The relational schema has 21 tables, but you will only need the 10 tables listed below.

Table name	No of rows
cast_info	39688662
char_name	3323337
company_name	252972
keyword	145970
kind_type	7
movie_companies	2915179
movie_keyword	4896658
name	4514054
role_type	12
title	2777526

The data model is the following. Notice that the arrows all represent a referring attribute in the origin table (named “..._id”). For instance the movie_keyword table has the attribute id and two additional attributes movie_id and keyword_id corresponding to the arrows pointing to title and keyword respectively.



Naming in the imdb2016 schema

The naming of tables and columns is given by the dump downloaded from IMDB and is unchanged (although somewhat confusing). E.g. references to the title, name and char_name tables are called movie_id, person_id and person_role_id respectively.

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 September 14).

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 |
+-----+
| actor |
| director |
+-----+
```

Responsive Applications, Web services and Databases

```

| 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 titles 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     |
+-----+

```

```

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

```

j) Find the title and production year of movies from "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 a character named "The Singing Kid"

title
Broken

1 row in set (0.05 sec)