University of Graz
Information modelling 1
Instructor: Dr. Sarah Lang
WS 2022/23

Exploring the diversity of *Taschen*'s film noir canon in SQL

Christina Dittmann
Christina.Dittmann@edu.uni-graz.at
Digital Humanities
Student number: 12229086
20 February 2023

Table of Contents

1. Introduction

For the final project in Information Modelling 1, I am modeling a list of the top 50 best films of the genre film noir. The list I base my model on stems from the publisher *Taschen*, which published a book describing various themes and motifs of film noir.[1] The original list serves to introduce film noir as a cinematic genre by compiling its most influential films. The purpose of the database is to analyze the distribution of aspects like the country in which the film was created or how many movies were created outside of the United States. In brief, this project aims at analyzing the diversity and complexity of *Taschen*'s canon.

Although film noir is not merely an American phenomenon – as German, Italian and other European filmmakers of the 20th century were of great influence –, most films that are in today's canon are American. But even outside of Europe and America film noir found recognition through Japanese films like Kurosawa's *Stray Dog* (1949) or Mexican films like Galindo's *The Devil's Money* (1953).

2. Stachowiak's Model Theory

In 1973, philosopher Herbert Stachowiak published a significant work on model theory called *Allgemeine Modelltheorie*. According to Stachowiak, models have three properties: mapping, reduction, and pragmatism.[2] He describes that mapping means that models are always models of something. And "something," here, may refer to mappings or representations of natural or artificial originals. The base for the model can even be models themselves.[3] Further, Stachowiak states that reduction describes that models do not have all attributes the original may have. Instead, models only convey certain information as they are reduced to the aspects that are important to the creator or user of the model. Pragmatism concerns the fixed functions a model has. The functions answer questions as to for whom, when or what the model is built.[4] It is thus not only important what original the model represents but the purpose thereof.

Stachowiak's model theory is used as a guideline for the model in this project. The model is meant to map *Taschen*'s canon included in their book. It reduces the given data to its essential so that it may answer the project's questions with the lowest level of redundancy. On the one hand, additional information is added for each film, like language and country, as *Taschen* included the film name, duration, director, screenplay, camera, music, and editing. On the other hand, screenplay, camera, music and editing are not used in this model. For this project, the

[1] The film noir canon can be found in Alain Silver et al.'s *Film Noir* (Taschen, 2017).
[2] Stachowiak, Herbert. *Allgemeine Modelltheorie.* Vienna: Springer (1973): 131.
[3] Ibid.
[4] Stachowiak. *Allgemeine Modelltheorie*. 132.

data needed to be filtered to what I deemed most important for the queries and what may convey information about diversity. Within this context, the film language or the country in which the film was shot seemed more significant than information as to who composed the music for the film. Apart from the purpose of extracting information about the data itself, the model also functions as an attempt to explore the possibilities of SQL and data retrieval from within databases.

3. Setting up the Database

The structure of the database was first defined in an ER diagram, which was then translated into a SQL database.[5] The dataset is scattered over different tables. The base table is called *films* and has the attributes film_id (primary key), film, director_id (foreign key), duration, and year. The table *films* includes 50 rows as the database captures 50 films. The films are sorted chronologically ranging from the years 1940 to 1960. Every film has its own id, represented by film_id – also ranging from 1 to 50.

Apart from *films*, there are also tables for *directors*, *languages*, and *countries*. The table *directors* (1:n cardinality) includes three columns, namely director_id (primary key), director, and gender. The column director makes up 13 rows. The table *languages* includes two columns called lang_id (primary key) and lang, short for language. Only four languages are registered under lang – English, French, German, and Italian. The table *countries* includes two columns called country_id (primary key) and country. The column country includes eight instances.

The database also includes auxiliary tables (n:m cardinality) as some films were filmed in different languages or countries. Due to the constraint of allowing only one value per cell in SQL, auxiliary tables are needed to solve the cardinality issue. The auxiliary tables created are *film_countries* and *film_lang*, which describe the countries in which the film was made and what languages are spoken in the film, respectively. The table *film_lang* includes three columns called film_lang_id (primary key), film_id (foreign key) and lang_id (foreign key). It has 51 rows. The table *film_countries* includes three columns called film_countries_id (primary key), film_id (foreign key) and country_id (foreign key). It has 53 rows. Every table in the database, including the auxiliary tables, therefore, has a primary key to uniquely identify a row. The auxiliary tables are not included in the ER diagram because they are mere aides to do queries.

The database could have been extended further by including screenplay, actors, or music. In fact, some screenplays were written by more than one writer and some directors wrote the screenplay themselves. In this case, it would have been helpful to create a table for all the

[5] The ER diagram can be found in the appendix.

persons included in the database to reduce redundancy. This table would then be linked to auxiliary tables by their primary key. This would be similar to the auxiliary tables *film_countries* and *film_lang*. However, for the present study and for reasons of simplicity, the information had to be limited to the existing tables.

4. Querying the Database

4.1 What is the average duration of the films?

```
SELECT AVG(duration)
FROM films;
```

The average duration of films is 101.18 minutes.

4.2 What is the shortest and what is the longest film?

```
SELECT film, MAX(duration) AS duration
FROM films;
SELECT film, MIN(duration) AS duration
FROM films;
```

The longest film is *Ossessione* with a duration of 140 minutes. The shortest film is *Detour* with a duration of 68 minutes.

4.3 What films are not (only) in English?

```
SELECT f.film, l.lang
FROM films f
JOIN film_lang fl ON fl.film_id = f.film_id
JOIN languages l ON l.lang_id = fl.lang_id
WHERE NOT fl.lang_id = 1
ORDER BY lang;
```

The non-English films in the database are *Les diaboliques* (lang = French), *Rififi* (French), *Shoot the Piano Player* (French), *The Third Man* (German), *It Happened in Broad Daylight* (German), and *Ossessione* (Italian).

4.4 Are there female directors in the database? What film(s) did they direct?

```
SELECT f.film, f.year, c.country, d.director
FROM films f
JOIN directors d ON d.director_id = f.director_id
JOIN film_countries fc ON fc.film_id = f.film_id
JOIN countries c ON c.country_id = fc.country_id
WHERE gender="female";
```

Carol Reed is the only female director in the database. She directed the 1949 film *The Third Man*, which was shot in Austria and the United Kingdom.

4.5 What is the percentage of non-English films?

```
WITH all_films AS (
SELECT COUNT(film) AS total
FROM films
),
non_eng_films AS (
SELECT COUNT(DISTINCT film) AS non_eng
FROM films f
JOIN film_lang fl ON fl.film_id = f.film_id
JOIN languages l ON l.lang_id = fl.lang_id
WHERE NOT fl.lang_id = 1
)
SELECT total, non_eng, (100.0 * non_eng / total) AS non_eng_percent
FROM non_eng_films, all_films;
```

The total includes 50 films, out of which six are non-English films, making up 12%.

4.6 What is the percentage of films shot outside of the United States?

```
WITH all_films AS (
SELECT COUNT(film) AS total
FROM films
),
non_us_films AS (
SELECT COUNT(DISTINCT film) AS non_us
FROM films f
JOIN film_countries fc ON fc.film_id = f.film_id
JOIN countries c ON c.country_id = fc.country_id
WHERE NOT fc.country_id = 1
)
SELECT total, non_us, (100.0 * non_us / total) AS non_us_percent
FROM non_us_films, all_films;
```

Eight out of the 50 films were shot outside of the United States, which equals 16%.

4.7 Are some directors represented with more than one film?

```
SELECT director, COUNT(film) AS count
FROM films f
JOIN directors d ON d.director_id = f.director_id
GROUP BY d.director_id
HAVING COUNT(*) > 1;
```

The query outputs the following directors: Alfred Hitchcock (count = 5), Raoul Walsh (2), John Huston (2), Robert Siodmak (2), Billy Wilder (2), Orson Welles (2), Nicholas Ray (2), Joseph H. Lewis (2), and Jules Dassin (2).

4.8 What is the average of films per director?

```
SELECT AVG(count) AS count_film_director
FROM (
SELECT COUNT(film) AS count
FROM films f
JOIN directors d ON d.director_id = f.director_id
GROUP BY d.director_id
);
```

The average of films per director is 1.31578947368421, or, if rounded, 1.32.

4.9 How many films represent each decade?

```
SELECT FLOOR(year/10)*10 AS decade,
COUNT(*) AS count_films
FROM films
GROUP BY decade;
```

Divided into decades, 28 films are from the 1940s, 19 films from the 1950s, and three movies from the 1960s.

5. Conclusion

This project shows that SQL is an effective way to analyze the diversity of a canon. Through different queries and commands, questions as to the percentage of foreign films or as to the average of film per director can be answered.

The queries show that the data based on *Taschen*'s canon has a low level of diversity as 16% are non-English films and 12% were shot outside of the United States. It must also be noted that the list only concerns Western films, thereby excluding films from South American or Asian countries. The data also includes a film by director Carol Reed, which gives merely a glimpse to the film noirs directed by women. The distribution of films per decade (ranging from 1940 to 1960) can be considered balanced as the 1940s birthed film noir, which is why the most important films were published in this decade.

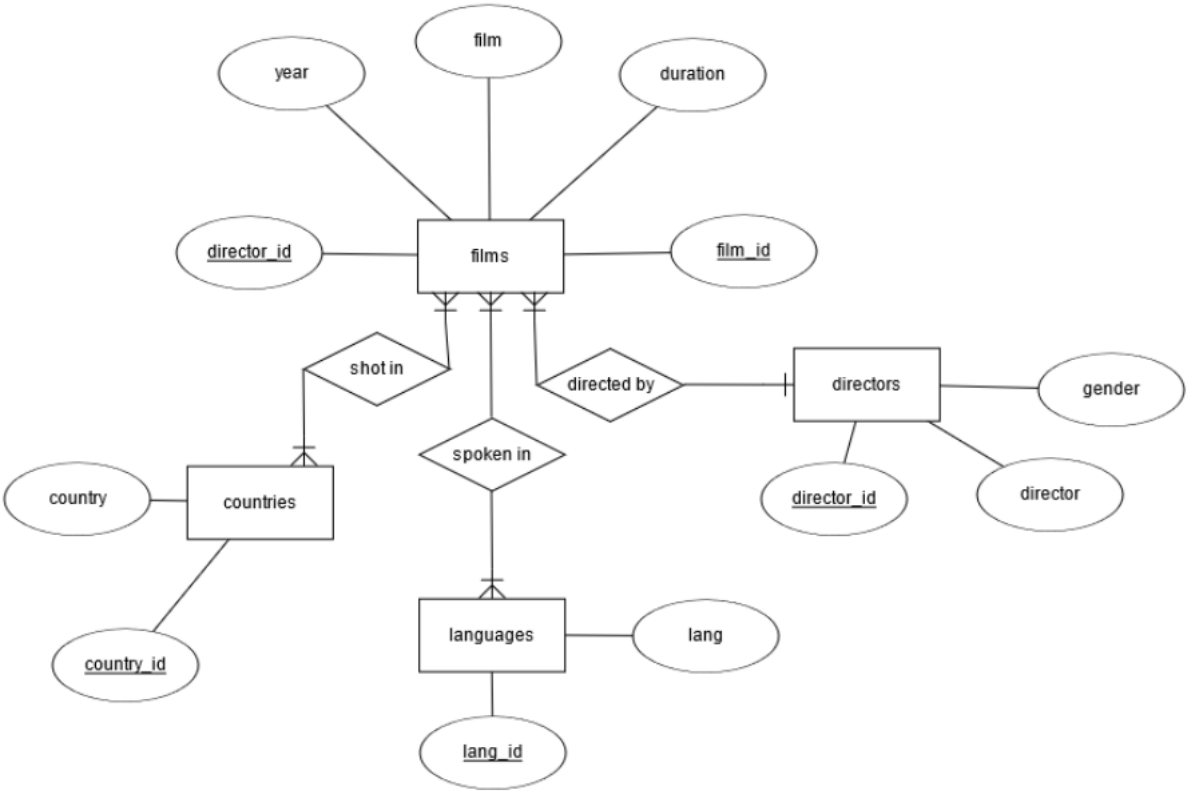
For a more in-depth analysis, the dataset could be extended to include more information about each film. Additionally, the data could be compared to other (film noir) canon compilations provided by agencies like the Internet Movie Database (IMDb).

Bibliography

Duncan, Paul, and Jürgen Müller, editors. *Film Noir*. Cologne: Taschen, 2017.

Stachowiak, Herbert. *Allgemeine Modelltheorie*. Vienna: Springer, 1973.

Appendix


film
year
duration
director_id
films
film_id
shot in
directed by
directors
gender
spoken in
country
countries
director_id
director
country_id
languages
lang
lang_id