

HS23 LTWA

Excercise 3:

SQL & Data Modeling

Notes on the submission:

- **Submission format:** zip-folder containing: a PDF or image file with your ER Diagram (task 2), a .sql file (task 3.1), a python script including your code for task 3.2 and 3.3 and a .txt file containing your name and your answers.
- Filename in the following format: gitlab-name_ltwa_exnumber, e.g. *max.muster_ltwa_ex3*
- Submit your zip file via the exercise module on OLAT if you want us to take a look at it. The module is only open until Wednesday, October 18, 12:00.
- The exercises are **not graded**.

If you have problems or questions just post in the OLAT forum or ask in the tutorials. If it is personal or urgent you can also contact us by e-mail. **Good luck!**

Language Technologies Web Application HS23
Department of Computational Linguistics

Lecturer: Johannes Graën
Tutors: Elina Stüssi, Lea Müller

1 Task: Getting familiar with SQL Commands

1. When working with language data, you often need to search for strings. Read the documentation for the LIKE operator:
 - a) LIKE:
<https://www.postgresql.org/docs/current/functions-matching.html#FUNCTIONS-LIKE>
2. The website <https://pgexercises.com/> provides interactive exercises for SQL queries. All of them are useful, but try to solve at least the following nine exercises:
 - a) Retrieve specific columns from a table:
<https://pgexercises.com/questions/basic/selectspecific.html>
 - b) Control which rows are retrieved:
<https://pgexercises.com/questions/basic/where.html>
 - c) Basic string searches:
<https://pgexercises.com/questions/basic/where3.html>
 - d) Removing duplicates, and ordering results:
<https://pgexercises.com/questions/basic/unique.html>
 - e) Join:
<https://pgexercises.com/questions/joins/simplejoin.html>
 - f) Count:
<https://pgexercises.com/questions/aggregates/count.html>
 - g) Aggregate count:
<https://pgexercises.com/questions/aggregates/count3.html>

- h) String search:
<https://pgexercises.com/questions/string/like.html>
- i) Case-insensitive search:
<https://pgexercises.com/questions/string/case.html>

2 Task: Data Modeling

The goal of the following two tasks is to create a database of Shakespeare's works and the characters that appear in those works.

1. Open the files `data/works.csv` and `data/characters.csv` and try to understand the data.
2. Design an Entity Relationship diagram based on the data. Make sure to follow the graphical conventions used in the lecture. You can draw it digitally or using pen and paper; please hand in a PDF or an image file.

Note: If you want to run your script multiple times for testing, it can make sense to add a statement to the beginning that automatically drops the table(s) if they already exist. Instead of dropping tables, a more elegant solution would be to check whether a certain table already exists and only create it if it does not exist.

3 Task: CREATE, CONNECT, INSERT and SELECT

3.1 Create the Tables

1. Connect to your database as owner and open a new console
2. Based on your ER diagram, define a schema for one or multiple tables and add the corresponding SQL Commands you used to the `.sql` file.
3. grant access to your `webapp_user` to allow necessary privileges on the created tables, add those SQL to your `.sql` file as well.

3.2 Connect to your database via Python

1. connect to your database as `webapp_user`, you can create a `.env` file again for your credentials. (To access those variables you might need to install the "python-dotenv" package and import the following:

```
from dotenv import load_dotenv
import os
```

you can call them like this:

```
host=os.getenv("DB_HOST")
```

2. In the same Python script, load the data from the CSV files into the table(s). For example, use the "csv" python library to read the files:

```
import csv
```

3.3 Answer some Questions about the Data:

At the bottom of the (same) .py script, run the queries necessary to answer the following questions:

1. Which characters say only one word?
2. In which play does a fairy named Mustardseed appear?
3. Do clowns appear more often in comedies or in tragedies?

Write your answers into a .txt file.