Chair for Algorithms
and Data Structures
Prof. Dr. Hannah Bast
Sebastian Walter

Information Retrieval
WS 2023/2024

http://ad-wiki.informatik.uni-freiburg.de/teaching

UNI
FREIBURG

Exercise Sheet 4

Submit until Tuesday, November 14 at **12:00 noon.**

Exercise 1 (10 points)

Extend *operations.py* from the code template on the Wiki as detailed in the following.

1. Implement the function *project* for selecting columns from a table.

2. Implement the function *select* for selecting rows from a table.

3. Implement the function *join* for joining two tables on one column. The join should be a hash-based equi-join as described in the lecture.

Exercise 2 (10 points)

First, download the TSV files containing the table data from the Wiki. Then, for the following query, determine three different sequences of operations (each operation in {*projection, selection, join*}) that all return the same result:

Actors who won a Golden Globe for a movie with an IMDb score of 8.0 or higher since 2010, including the name of the movie, the role played, and the type of Golden Globe award.

Enter your sequences of operations as code into *queries.py* using your implementations from exercise 1. You can and should look at the corresponding SQL statement in *queries.sql* for guidance. We also provide an example for a different query and SQL statement in *queries.py* which can be executed using *python3 queries.py *.tsv --example*, assuming you downloaded the TSV files to the same directory.

Verify the correctness of your results by executing *python3 queries.py *.tsv* and comparing its output to the one of SQLite3. Execute *make sql* or *cat queries.sql | sqlite3* to get the SQLite3 output.

Enter the runtimes of your three sequences of operations and the runtime of SQLite3 with and without indices into the table on the Wiki, following the format of the rows already there.

Commit your code to our SVN, in a new subfolder *sheet-04*. The TSV files should not be committed.

As usual, in your *experiences.md*, provide a brief account of your experience with this sheet and the corresponding lecture. Make sure to add a statement asking for feedback. In this statement specify to which degree and on which parts of the sheet you want feedback. In addition, say how much time you invested and if you had major problems, and if yes, where.