CPSC 531 - Advanced Databases - Fall 2019

Project 2, due November 5

RDBMS and Database

The RDBMS for this project is a Python program that you and your team will implement.

The database is Luis Rocha's Chinook Database, modified for use with Python.

Platforms

You may use any platform to develop and test your code, but note that per the Syllabus the test environment for projects in this course is a Tuffix 2019 Edition r2 Virtual Machine with Python 3.6.8. It is your team's responsibility to ensure that your code runs on this platform.

Loading the Database

Download the database file `Chinook_Python.py` and place it in directory where you will develop your Python code. The database can be loaded with the following command:

```
from Chinook_Python import *
```

You will find variables named `Artist`, `Album`, `Customer`, etc. containing sets of `namedtuple` objects.

Note: an earlier version of `Chinook_Python.py` did not export the `InvoiceLine` relation by default. You can import it explicitly (e.g., `from Chinoook_Python import InvoiceLine`), or use the link above to download an updated version.

Relational Operators

Implement the following functions in Python:

- `select(relation, predicate)`
- `project(relation, columns)`
- `rename(relation, new_columns=None, new_relation=None)`
- `cross(relation1, relation2)`
- `theta_join(relation1, relation2, predicate)`

The predicate for `select()` should be a function that takes a single `namedtuple` as an argument and returns `True` or `False`.

The predicate for `theta_join()` should take two `namedtuples` and return a `bool`.

The `new_columns` and `new_relation` parameters to `rename()` are optional. if neither argument is provided, return the original `relation`.

Extra Credit

Implement `natural_join(relation1, relation2)`.

Queries

The file [`queries.py`]() contains four variations of the first query from [Project 1]():

1. Combining σ and × to implement θ-join
2. Performing σ after θ-join
3. Performing σ before θ-join
4. Natural join (Run this If you did the extra credit.)

All of the queries above should return the following set:

```
{Result(Title='Blood Sugar Sex Magik'),
 Result(Title='By The Way'),
 Result(Title='Californication')}
```

When the relational operators are implemented and the queries above work correctly, write code to run the last query from Project 1.

Performance Measurement

As a rough approximation of the processing required for each query, instrument your functions to measure the cardinality of the result set for each relational operator. When a query completes, print the total number of tuples returned during processing.

1. What do you observe about the queries listed above?

2. Can you rewrite the last query from Project 1 to minimize the number of tuples processed?

Python Tips

If you are new to Python, see [A Whirlwind Tour of Python]().

You may find the following useful:

- Lambda expressions can be used to write predicates succinctly.
- Named tuples include a `._fields` attribute.
- Use `next(iter(relation))` to retrieve the first named tuple in a relation.
- Use `tuple.attribute` to access fields of named tuples directly.
- Use `getattr(tuple, attribute_name)` to access a field of a named tuple given the name of the attribute as a string.
- A list of arguments can be unpacked for a function call requiring separate positional arguments.
- Set comprehensions can use a syntax similar to list comprehensions to construct sets concisely without loops.

Submission

Submit your project by uploading your `.py` files and any other relevant artifacts to the `project2/` subdirectory of the folder that was shared with you on Dropbox.

You may work alone, or make a single submission for a team of 2-3 students. If you work in a team, only one submission is required, but for safety consider uploading copies to each team member's submission folder. (Make certain, however, that the copies are the same in each case; the instructor will not attempt to ascertain which is the "real" submission.)

A printed submission sheet will be provided on the due date. To finalize your submission, fill out the sheet with the requested information and hand it in to the professor by the end of class. Failure to follow any submission instructions exactly will incur a **10%** penalty on the project grade for all team members.