Kelly Breedlove
Dr. VanDrunen
Project 6
1 May 2015

Design:

- Driver.py
  - If name is main: a command line interface into the graph for custom queries, run the program from here
    - NOTE: the printed graph is captured in the text file graph for reference
  - This class also contains a function that uses DatabaseUtils.py to gather data and make a list of vertices to be graphed
- DatabaseUtils.py
  - If name is main: pass / temporary testing
  - This class contains functions that use SQL statements to retrieve necessary data from the databases
- Graph.py
  - If name is main: none
  - This contains the class Vertex and the class Graph
- runTests.py
  - The test driver for the following unittests:
    - TestArtistSimilarityGraph.py
      - NOTE: this test runs for about a minute
    - TestDatabaseUtils.py
    - TestGraph.py

Question:

- What music artists are similar to a given artist? How removed are some artists from others?
  - These questions were prompted by the popularization of Internet radio that generates playlists of similar music based on a single artist or album.

Data:

The dataset used is called Million Song Subset. It is an sqlite3 database that includes data on artists, their songs, their albums, and similar artists. The specific tables used are formatted as follows:

```
subset_track_metadata.db
CREATE TABLE songs (track_id text PRIMARY KEY, title text, song_id text,
release text, artist_id text, artist_mbid text, artist_name text, duration
real, artist_familiarity real, artist_hotttnesss real, year int);

subset_artist_similarity.db
CREATE TABLE artists (artist_id text PRIMARY KEY);
CREATE TABLE similarity (target text, similar text, FOREIGN KEY(target)
REFERENCES artists(artist_id), FOREIGN KEY(similar) REFERENCES
artists(artist_id) );
```

Because the databases are in sqlite3, I chose to write the code in python for an easy interface between languages. In this way, I wrote a class of utility functions, DatabaseUtils.py, to make the necessary queries of the database.

Graph Implementation:

Vertex and Graph are implemented using the adjacency list approach. I chose this since iterating over adjacency is important for the way this data relates.

The class functions for Vertex and Graph were chosen based on what was necessary to best answer the questions asked of the data. Writing Vertex and Graph in python were a little tricky in some situations because of duck typing and the lack of encapsulation. However, iterating over things is much easier in python than Java.

Algorithm:

The most interesting algorithm used in graph is Depth First Search. A search is performed starting at a given artist in order to find out how relates to the other artists. The code for that part is nearly identical to the code given in class. Once the paths are populated, the paths dictionary is used to backtrack from the second artist given to find the path between it and the artist that was the origin.

Results:

The driver is written so that you can make new queries into the graph to learn about whichever artists interest you. Here are some example results to my original questions with The Beatles as the primary artist in question.

- Similar artists for The Beatles include:
    - Big One
    - Bob Dylan
    - David Bowie
    - Electric Light Orchestra
    - Flamin' Groovies
    - Jimi Hendrix
    - Led Zeppelin
    - Marshall Crenshaw
    - Teenage Fanclub
    - The Animals_ Sonny Boy Williamson
    - The Moody Blues
    - The Rolling Stones
- The path from The Beatles to Nirvana is:
    - The Beatles
    - The Rolling Stones
    - Ultimate Tracks
    - Tom Petty And The Heartbreakers
    - Wailing Souls
    - The Twinkle Brothers
    - Tony Tuff
    - Yami Bolo
    - Maxi Priest
    - Singing Melody
    - Tyrone Taylor
    - Winston Reedy
    - Peter Spence
    - Rising Lion
    - Super Cat
    - Major Lazer / Vybz Kartel / Afrojack
    - The Bug Featuring Ricky Ranking
    - iTAL tEK
    - Pariah
    - Photek
    - John Digweed
    - Way Out West
    - Wamdue Project

- Timo Maas
- The Prodigy
- The Crystal Method
- Thomas Dolby
- Wang Chung
- The Police
- Wyclef Jean
- Usher Featuring The Nu Beginning
- Toni Braxton
- Sade
- Simply Red
- Wet Wet Wet
- Mike And The Mechanics
- Eddie Money
- The Ronettes
- The Shirelles
- Johnny Nash
- Jackie Wilson
- Tyrone Davis
- Z.Z. Hill
- The Robert Cray Band
- Walter Trout
- Sue Foley
- The Vaughan Brothers
- Taste
- The Edgar Broughton Band
- The Sensational Alex Harvey Band
- Uriah Heep
- Whitesnake
- Warrant
- Vixen
- Evanescence
- The Smashing Pumpkins
- Nirvana

I personally find the path between artists to be the most interesting part of the results. In my opinion, the data on similar artists, particularly for bands from the early 60’s, are weak. It seems to be based largely on decade rather than sound. I would be interested in running this program with the full Million Song Database rather than just the Million Song Subset.