INLS 560 – Final Project, Assignment 7: Review Dashboard

Date Assigned: Fri. Nov 7, 2014

Completion Date: Fri. Dec 6, 2014 (8:00 am)

There are several milestone due dates. They are as follows:

Due Dates

Nov. 14 (midnight) Options: 1, 5, 6

- submit your .py files and your database

Nov. 21 (midnight) Options 2-4

- submit your .py files and your database

Dec. 6 (8:00 am)

- You will demo your application
- You will also show your code and explain your design decisions (such as the functions you created, how your code works together)

In this assignment, you will write code and during our exam time, demo your application and explain your design decisions. Each student will have six minutes to demo their application and explain their design decisions.

Software issues:
If you feel there are mistakes in this assignment, check the web page and Sakai for corrections, and

report them to us if they have not been made.

Description

Many business owners often wonder what customers think of their business. One way to find the

answer to this question is to send users surveys. An alternative is to read customer reviews from

websites like Yelp. Both options require business owners to read through many customer reviews. To

help business owners, you will create an application that shows the owner:

- The number of positive and negative reviews.

- The percentage of positive and negative reviews.

- The top n common phrases used in reviews (where n is a number less than 20)

- The top n common phrases for positive reviews (where n is a number less than 20)

- The top n common phrases for negative reviews (where n is a number less than 20)

- n examples of positive reviews (where n is less than 6)

- n examples of negative reviews (where n is less than 6)

Requirements

This program requires you to:

- Create sqlite tables
- Use The sqlite3 built-in library
- Loops
- Database
- Objects
- APIs
- Dictionaries

Assignment Specification

Download the review database from (link):
http://ils.unc.edu/courses/2014_fall/inls560_001/assignments/INLS_560_Final_Project.zip

This is a huge file as it contains the review database along with other code and data files.

Tables

- Create a table named **sentiment** that holds the sentiment (positive, negative) for each review
    - The columns should be review_id, business_id, sentiment
- Create a table named **review_stats** that holds the number of positive reviews, negative reviews and the percentage of positive reviews, and negative reviews for each business
    - The columns should be business_id, number_of_positive_reviews , number_of_negative_reviews, percentage_of_positive_reviews, percentage_of_negative_reviews,
- Create a table, **common_phrases** that holds the common phrases
    - phrase_id, business_id, common_phrase, frequency_of_phrase

Classes and Objects

At a minimum, there should be a class for each table.

Sentiment (positive, negative reviews)

To determine the sentiment of reviews we will use the TextBlob API
(http://textblob.readthedocs.org/en/dev/). Before you use this api, at the command line, you must type:

(Windows) python.exe -m textblob.download_corpora

(Mac) python -m textblob.download_corpora

To use this API:

from textblob import TextBlob

from textblob.sentiments import NaiveBayesAnalyzer

SENTIMENT_TYPE = 0

POSITIVE_SENTIMENT_VALUE = 1

def main():

   blob = TextBlob("I love this library", analyzer=NaiveBayesAnalyzer())

   sentiment = blob.sentiment

   print "This word has a", sentiment[SENTIMENT_TYPE], "sentiment the positive value is", sentiment[POSITIVE_SENTIMENT_VALUE],\

main()

User Interface

Please enter the name of a business: Business A

Please choose an option:
1. Show the number of positive and negative reviews, the percentage of positive and negative reviews
2. Show the top n common phrases
3. Show the top n common phrases for positive reviews
4. Show the top n common phrases for negative reviews
5. Show examples of positive reviews

6. Show examples of negative reviews
7. Exit

Other Code

Write code that reads information from the **review** table, determines the sentiment, and inserts information into the **sentiment** table.

Write code that reads information from the **sentiment** table, computes the # of positive and negative reviews. It should also compute the percentage of positive and negative reviews. This information should be inserted into the **review_stats** table.

Write code that reads information from the **review** table, computes the frequency of each word in review, insert words with a frequency of n into the **common_phrases** table.

Write code that reads from both the **sentiment** and **common phrases** tables to get the common phrase for each sentiment (positive and negative).

Grading

Programs will be graded based on whether they display the correct output, the correct logic, and style.

In this assignment style means, make variable names meaningful. Do not create one letter variable names or variable names that do not have anything to with the assignment. The program must not only print the correct values, but the code must actually perform the correct operations.

Getting Help

If you have trouble, please post a question on Piazza before contacting me. Before posing a question, please check if this question has been asked before. This will reduce post clutter and reduce our burden. Repeat questions will be ignored by the instructors.

Piazza allows anyone to respond. So if you see a question that you think you can respond to, please do so, as that will reduce our burden and help you "teach" your fellow students.

Positive luck!