CSE 482 Exercise 3 (Due date: October 02, 2020)

The purpose of this exercise is to help you get started using MySQL and MongoDB databases. Follow the instructions below to complete the exercise. Save your IPython notebook as exercise3.ipynb.

1. You will use the wiki_edit.txt file from Exercise 1. Make sure the data file is in the directory where you will be writing your python code.
2. Write a python script using the mysql-connector library to performs the following steps:
   a. You need to install MySQL and PhpmyAdmin on your own machine. The instructions are posted in https://piazza.com/class/ked1akrwvs45xt?cid=53
   b. Open a connection to the database server on your machine.
   c. Create a table named Wiki_Edit, which contains the following schema:
      > RevisionID: int Primary Key
      > ArticleName: varchar(500)
      > EditDate: date
      > UserName: varchar(50)
   d. Read the file wiki_edit.txt and parse each line to identify the individual fields. For this exercise, do not use the read_table and DataFrame approach shown in lecture 4. Instead, you should read the file directly, parse each line, and store the fields from each line into the database. You should refer to the example given in slide #16 in lecture 2 on how to read from the data file directly and split each line into the individual fields.
   e. Query the database to find the Article name with largest number of edits.
3. Install MongoDB to your machine.
   a. Launch the MongoDB server.
   b. Using pymongo, write an IPython notebook script that will store the entries in the wiki_edit.txt file to a collection named wiki in MongoDB. To store the data, you need to convert each line into a dictionary object. The keys to the dictionary correspond to RevisionID, ArticleName, EditDate, and UserName and the values correspond to the individual entries in each line of the file:
      > columns = line.strip().split(' ')
      > record = { 'RevisionID': columns[0], …., 'UserName': columns[3]}
      > db.wiki.insert_one(record)
   c. Write a query to find the date and username of editors who edited the "Anarchism" article:
      > revisions = db.wiki
      > for line in revisions.find({"ArticleName": "Anarchism"}):
      >   print line["EditDate"], line["UserName"]

Deliverable: Submit the IPython notebook exercise3.ipynb that contains all your code.