Open Source Software — CSCI-4966-01 — Spring 2019
Test 2
April 16, 2019

Name: _____

RCS ID: | | | | | | | | | @rpi.edu |

RIN#: _____

Honor pledge: On my honor I have neither given nor received aid on this exam.

Please sign here to indicate that you agree with the honor pledge: _____

Instructions:

- You have 90 minutes to complete this test.
- Clearly print your name, RCS ID (in all caps.) and your RIN at the top of your exam.
- This test is open book, open notes and open computer. You **m**ay not use the internet. Please turn off your wifi.
- There are **7 questions** on this test worth a total of **120 points**.

1. Short answers. Feel free to use your notes. (40 pts)

   (a) There are numerous fields in which Scientific Computing is essential. In class, we defined 16. Give 4 of the fields we called out. (8 pts)

      i.

      ii.

      iii.

      iv.

   (b) Looking at the Angry Birds Game, answer the following two questions. (4 pts)

      i. Which module encapsulates the physics of the simulation?:

      ii. The entirety of the physics simulation for a given time step is contained in a single call. What is the call to advance the simulation one step?:

   (c) Name 3 statistical packages with a BSD based license (6 points):

      i.

      ii.

      iii.

(d) Name four benefits of incremental testing (8 points):

i.

ii.

iii.

iv.

(e) What percent of open source projects have a single developer (as of 2013). Any answer within 5 points of the correct solution will be accepted. (2 points)]?

i.

(f) Name three open source governance models (6 points):

i.

ii.

iii.

(g) Who `"owns"` an open source project? (4 points)

i. (Who:)

ii. (Why:)

(h) What is the license for TensorFlow? (2 points)

2. *Scientific Computation* Feel free to review, edit or run code from the Scientific Computation lecture or lab to answer the following questions. (16 pts):

   Note: If you import `networkx` into python and issue the command `help(networkx.shortest_path)` it will provide you with additional information on the shortest path algorithm. In particular, notice that if you do not provide a target, then the algorithm returns a dictionary where the keys are words that can be reached from the source and the values are the list of nodes. This information will help you with this question.

   (a) Consider your word ladder code to find the shortest path from one five letter word to another.

      i. Assuming that you are not allowed to change the order of the letters, what is the length of the longest, shortest path from the word *party*? (4 points)

      ii. What is the final word in the path? (4 points)

   (b) Now consider the degree of a word as the number of other words that can be made from it by changing a single letter and keeping the word order the same.

      i. What words have the maximum degree of 25? (4 points)

      ii. There are far more words with degree 0. How many are there? (4 points)

3. *Statistical Computation* Feel free to review, edit or run code from the Statistical Computation lecture or lab to answer the following questions. (10 pts):

   (a) Consider the **topmovie** data shown below.

	rank	**name**	**box**	**date**	**year**
1	1	Avatar	759.563	December 18	2009
2	2	Titanic	600.788	December 19	1997
3	3	The Dark Knight	533.184	July 18	2008
4	4	Star Wars: Episode IV - A New Hope	460.998	May 25	1977
5	5	Shrek 2	437.212	May 19	2004
6	6	E.T. the Extra-Terrestrial	434.975	June 11	1982
7	7	Star Wars: Episode I - The Phantom Menace	431.088	May 19	1999
8	8	Pirates of the Caribbean: Dead Man's Chest	423.416	July 7	2006
9	9	Toy Story 3	414.638	June 18	2010
10	10	Spider-Man	407.681	May 03	2002
11	11	Transformers: Revenge of the Fallen	402.077	June 24	2009
12	12	Star Wars: Episode III - Revenge of the Sith	380.271	May 19	2005
13	13	The Lord of The Rings: The Return of The King	377.019	December 17	2003

   i. Give a command or commands to get the Minimum, 1st Quartile, Median, Mean, 3rd Quartile, and Maximum values for the box office. (4 points)

   ii. If we want to look at the relationship between the box office and the year, how would we generate a scatterplot with year on the *X-axis* and box office on the *Y-axis*? Type the command below: (6 points)

4. `Testing and Continuous Integration` Feel free to review, edit or run code from the Testing and Continuous Integration lecture or lab to answer the following questions. (15 pts):

   Consider the following Python module implementing the merge_sort algorithm:

```
import random

def merge(L1, L2):
  """
  Assume L1 and L2 are sorted.
  Create a new list L that is the merged
  version of L1&L2.
  """
  L = []
  i = 0
  j = 0
  while i < len(L1) and j < len(L2):
    if L1[i] < L2[j]:
      val = L1[i]
      L.append( val )
      i += 1
    else:
      val = L2[j]
      L.append( val )
      j += 1
  ## at this point, either L1 or L2 has run out of values
  ## add all the remaining values to the end of L.
  L.extend(L1[i:])
  L.extend(L2[j:])
  return L

def merge_sort_recursive(L):
  """
  Complexity: O(n logn)
  The function calls itself recursively logn times,
  and each time about n elements are merged.
  """
  if len(L) <= 1:
    return L

  length = len(L)
  mid = length // 2
  left = merge_sort_recursive(L[:mid])
  right = merge_sort_recursive(L[mid:])
  return merge(left,  right)

if __name__ == "__main__":
  ##Testing code
  k = 10
  L = list(range(k))
  random.shuffle(L)
  print("Before:", L)
  L = merge_sort_recursive(L)
  print("After:", L)
```

Generate a python file, *test_merge.py* that uses the unittest framework to thoroughly test the *merge* and *merge_sort_recursive* functions. Assume *white box* testing. You should be able to come up with at least 3 test cases for *merge* and at least 2 test cases for *merge_sort_recursive* Write your *test_merge.py* function in the space below:

5. `MongoDB` Feel free to review, edit or run code from the MongoDB lecture or lab to answer the following questions. (15 pts):

   Consider the definitions file we used for our MongoDB lab, particularly for *checkpoint4.py* and *checkpoint5.py*.

   (a) Write the sequence of commands to reset the database to the contents of the file `definitions.json`. Use `$` to indicate commands typed on the command line and `>` to indicate commands typed into the mongo prompt. You should assume that the database is already in mongodb with the name `mongo_db_lab`, i.e. (3 points)

```
$ mongo
MongoDB shell version v4.0.0
connecting to: mongodb://127.0.0.1:27017
MongoDB server version: 4.0.0
> show dbs
admin         0.000GB
config        0.000GB
local         0.000GB
mongo_db_lab  0.000GB
> quit()
$
```

   Write your commands in the space below:

(b) How many entries in the database have the string `" RPI "` in them? Feel free to run a query to find out (2 points)

(c) Now write Python code to search through the database, replace every occurrence of `" RPI "` in a definition with `" Rensselaer Polytechnic Institute "`, and write the modified records back out to the database. Be careful with the search and replacement strings. You do not want to change the string `"(RPI)"`. Make sure you keep all other fields (*i.e.* the "word" field) correct. (Also, if you are testing your code make sure that you reset your database before running.) (10 points) Write your code in the space below:

6. `Virtualization and Containers` Feel free to review, edit or run code from the Virtualization and Containers lecture or lab to answer the following questions. (10 pts):

   (a) Look at the following command to run a docker instance.

   ```
   $ docker run -i -t -p 8888:7777 ubuntu:latest
   ```

   Briefly describe what the following parts of the line do to the execution: (5 points)

   i. -i:

   ii. -t:

   iii. -p 8888:7777

   iv. ubuntu

   v. :latest

   (b) Assume you have the Dockerfile below:

   ```
   # Comments in Dockerfiles
   FROM ubuntu:latest

   RUN apt-get update
   RUN apt-get install sudo
   RUN apt-get --yes install apt-transport-https
   RUN apt-get --yes install dbus
   RUN apt-get --yes install python3
   RUN apt-get install python3-pip

   RUN mkdir -p /data/db

   RUN useradd -d /home/ubuntu -ms /bin/bash -g root -G sudo ubuntu
   RUN echo "root:Docker!" | chpasswd
   RUN echo "ubuntu:ubuntu" | chpasswd

   USER ubuntu
   WORKDIR /home/ubuntu
   ```

   i. Write the command to compile this docker file into an image with the name profsfavorites (2 points):

   ii. What commands would be added to the Dockerfile to install `git` and `vim` (both are available through apt-get) (2 points)?

   iii. How many user accounts are there on the machine (do not count the root account) (1 point)?

7. `TensorFlow` Feel free to review, edit or run code from the TensorFlow lecture or lab to answer the following questions. (14 pts):

   (a) Starting from a color image, what 5 steps did we need to do to curate our data to match the starting data for the NIST fashion data? (2 points each)

   i.

   ii.

   iii.

   iv.

   v.

   (b) Assume you have a trained model in variable "m", a test data set in variable "t" and labels for "t" in "l", write code to evaluate the model on the test data and print out the loss and accuracy data for the test data set. (4 points)