ML110 Homework 6

MLLO 20220524

Submission

- Deadline: **6/10 Fri. 23:59**
- Submission: **NTU COOL**
- File name & format
  - <學號>.zip, ex. r09900001.zip
  - (unzip 後須有以下檔案，且不能有其他檔案)
    - mongo.txt
    - neo4j.txt
    - report.pdf
- 檔案格式錯誤一律扣**10%**
- 說明:
  - mongo.txt / neo4j.txt : 各題 mongoDB (part I.) / neo4j (part II. & part III.) 的 code
  - report.pdf : 各題的code 截圖 & results，煩請題號標示清楚。
- Version requirements
  - MongoDB v3
  - Neo4j v4

Homework 6 - Forewords: Reminding

- To complete this homework, you need to have downloaded the MongoDB server and related client-side programs, and the Neo4j database server and the related client-side programs. The related information has been presented to you in our course notes. You can also find all necessary information from MongoDB and Neo4j's documentation.
- You also need to have practiced and gotten familiar with these technologies, and gotten familiar with how to write MongoDB queries, and the Cypher queries.

Homework 6 (2)

- **Part I (MongoDB)**
- **Task 1**
  Load the hw6_student_list.csv [link] file into collection “students” in database “hw6” in MongoDB, and write a MongoDB query to return the information (the document) about yourself. Note that we did not cover how to load CSV file in our lectures. You are responsible to find out how yourself. (Hint: download and install the MongoDB Database Tools package, and use the mongoimport) **(5%)**
- **Task 2**: Write a MongoDB query to return the information about you and your peer students. **(10%)**
- Note: The definition of “peer student” is the same in the SQL homework. Copied here for your reference:
  - For undergraduate students, your peer students are those in the same department and same year as you.
  - For graduate students, your peer students are those in the same graduate program and same year as your.
  - peer students 不包含自己。若沒有同系且同年級，就 output empty 並請附註說明
- **Task 3**: For each “系級" in the “students” collection, find out the number of students in it. Write a MongoDB query to return this information (Hint: aggregation. **10%**)

Homework 6 (3)

- **Task 4**: For the documents for each student, add a new field “加入日期", and set the “2022-03-01”. Then return the information about your and your peer students again to make sure your update is successful. **(10%)**
- **Task 5**: Add the following students into your “students” collection.

加入日期	身份	系所	年級	學號	姓名
2022-06-02	旁聽生	歷史系	1	b09900201	小花
2022-06-02	校內生	歷史系	4	b06900332	小草
2022-06-02	校內生	機械系	4	b06502055	小天

  - Write a query to return yourself and these new students to prove you have successfully inputted the data. **(5%)**
  - Note: You are free to choice to input the date using string “2022-03-01”, or as the ISODate() object.

- **Task 6**: (Challenge problem)
Design a increment aggregation pipeline to calculate the number of student for each “系級", and store your result in a “tally” document in your “students” collection. Run your query with date as to “2022-03-31” first. Print out the “tally” document. Then run your query again with date set to “2022-06-10”, and print our the “tally” document. **(10%)**
  - Hint: you shall need to learn the 2nd half of MongoDB course note well, especially the last three pages.
  - Reference:
```
{ $dateFromString: {
      dateString: "2017-02-08"
} }
```
  - Reference:
```
ISODate("2017-02-08T00:00:00Z")
```

Homework 6 (4)

- **Part II (Graph DB Basic)**
- **Task 1**: Load the hw6_student_list.csv (the same as Part1) into Neo4J graph database and create one node for each students. **(5%)**
- **Task 2**: Create a “peer” relationship between you and each of your peer students. **(10%)**
- **Task 3**: Write a Cypher query to return you, and your peer students in a list **(10%)**

- **Part III (Graph DB Advanced)**
- **Bonus: Fill out your 5 hobbies in the file** hw6_hobbies
  before 5/26 Thu. 23:59 (5%)
- **Task 1**: Load the hw6_hobbies.csv into the database and create the “necessary nodes and relationships” (Think carefully how you want to represent hobbies in your GDB.) **(5%)**
- Now assume you want to expand your hobbies. Logically, you need to first find out all the students in this class who has the at least one common hobbies as you. Let’s call these your “**hobby friends**”.
- The second step is to find other students who have at least one common hobby with your hobby friends, let’s call them “**foaf**” (**friend of a friend**). The hobbies of these people are the potential new hobbies you want to acquire.

Homework 6 (5)

- **Task 2**: Write a cypher query to print out your “foaf” and their associate new hobbies. **(10%)**
  - You yourself should not be in this list. If you yourself appear in this → -4%
  - Your hobbies friends’ hobbies should not appear in the hobbies of your foaf. If they do → -4%

- **Task 3**: Write a cypher query to print our all the new hobbies of your foaf. **(10%)**
  - Your old hobbies should not appear in the list. If they do → -4%
  - There should not be duplicates. If they do → -4%

Homework 6 (6)

- **Part III 2 (Graph DB Advanced) 的部分，report 內容除了包含你的結果以外，還要有簡單的說明/解釋**
- **Neo4j 介面上可以將matching到的subgraph 畫出來，並輸出 .jpg/.png**
- **輸出結果介面中，右上角有下載符號**