EE5178
111-2 Homework 6

20230531

Homework 6 - Reminding

- To complete this homework, you would need download the related MongoDB and Neo4j servers and client-side programs. Alternatively, you access the cloud versions of the software. An excerpt of related information can be found in our course notes. You can also find all necessary information from the documentation in their respective websites.
- You also need to have gotten familiar with these technologies
- You are highly encouraged to have practiced and gotten familiar with how to write MongoDB and Cypher queries, before you embark on completing the homework.

Homework 6 (2)

- Part I (MongoDB)
- Task 1: Load the EE5178 student CSV 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) (10%)
- Task 2: Write a MongoDB query to return the information about you and the students in your group. (10%)
- 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 “join_date", and set the “2023-03-01”. Then return the information about you and your group members again to make sure your update is successful. (10%)
- Task 5: Add new students into your “students” collection, using the “new_student_list.csv” file provided.
  Write a query to return yourself and these new students to prove you have successfully inputted the data. (5%)
  Note: You are free to choose to input the date using string, 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 “2023-03-31” first. Print out the “tally” document. Then run your query again with date set to “2023-06-10”, and print out the “tally” document. (5%)
  - 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 EE5178 student CSV into Neo4J graph database and create one node for each students. (5%)
- Task 2: Modify the database so that you and the students in your project group are recorded as in the same group. (Hint: there is more than one way to do this. Think and design carefully how you want to do it.)
- Task 3: Write a Cypher query to return you, and students in your group in a list (10%)

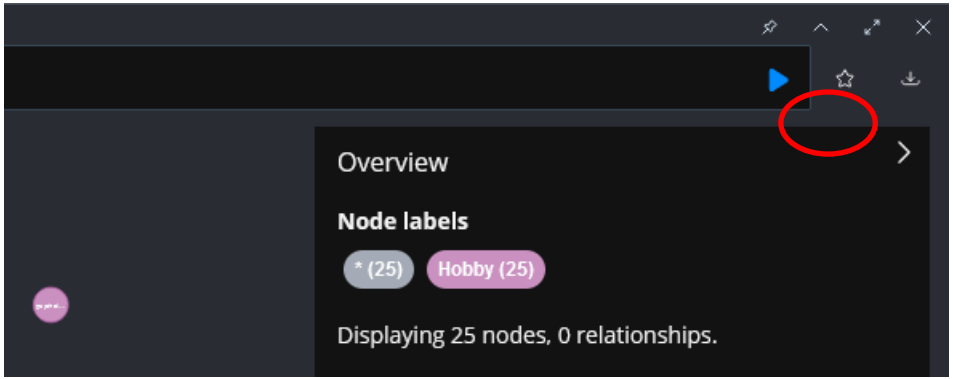
- Part III (Graph DB Advanced)
- Task 1: Load the student hobbies CSV into the database and create the necessary nodes, relationships, and/or properties so that hobby information is recorded into the database (Think carefully how you want to represent hobbies in your GDB.) (5%)
- Task 2: Now assume you 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".
  Write a cypher query to print out your "hobby friends" and their associate new hobbies. (10%)
  - Your old hobbies should not appear in the list. If they do → -4%

Homework 6 (5)

- Task 3: (Challenging) The second step is to find other students who have at least one common hobby with your hobby friends, let's call them "foaf" (friends of a friend). The hobbies of these people are the potential new hobbies you want to acquire.
  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 hobby friends / hobbies should not appear in the foaf / hobbies of your foaf. If they do → -3%
  - There should not be duplicated. If there is → -3%
- Task 4: (Challenging) Actually you should treat everyone in your project group as good friends. Now, the hobby friends of your group members can also become your friends. Let's call them "foaf2" (friends of a friend 2)
  Write a cypher query to print your "foaf2". (5%)
  - There should not be duplicates. If there is → -4%
- Task 5: You want to find out what new friends you can acquire from your project group connections. Write a Cypher query to list all your "foaf2" excluding "foaf". That is, list all your "foaf2" who is not in your "foaf" (5%) (In other words, list all people who is a hobby friend of some of your group members, but who is not your friend of a friend defined in task 3).

About Part III

- 沒有填寫 hobbies 或是和別人的 hobbies 完全沒有重疊的同學，請選擇一位有填寫的同學當成自己，以完成相關 task，並在report 中註明選擇的同學。
- 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



Version requirements

- MongoDB 6
- Neo4j Community Edition 5.8.0

Homework 6 submission

- Deadline : 6/15 Thur. 23:59 (GMT +8)
- Submission : NTU COOL
- File name & format
  - 上傳檔案: hw6_{id}.zip, e.g. **hw6_r09900001.zip**
  - 解壓縮後需包含
    - 最外層資料夾: hw6_{id}, e.g. **hw6_r09900001**
    - hw6_mongo.txt / hw6_neo4j.txt
      - 各題 mongoDB (part I.) / neo4j (part II. & part III.) 的 code
    - Report: hw6_{id}.pdf, e.g. **hw6_r09900001.pdf**
      - **各題的code 截圖 & results，煩請題號標示清楚。**
    - 沒有最外層資料夾、檔名錯誤一律扣10分

Homework 6 submission

- Delay
  - One day: original score * 0.8
  - More than two days: get no points
- TA hour: Mon. 09:00 – 11:00 @ 博理603
- TA mail: ntudbms2023.ta@gmail.com
- Q&A : NTU COOL 討論區 || TA mail