Milestone 2

Computer Science 316

11/3/2018

Eddie Yang, Frank Qu, Jack Claar, Yifan Yin

Progress Report:

**For our milestone 2, we are changing the project topic and most of the information from milestone 1, so we are rewriting the information for milestone1 in this report .**

Our Application:

There are so many classes offered per semester, and it sometimes becomes a tedious task for students to navigate through the list of choices and to find the courses of their interest. This database helps students to find classes with specific professors, areas of knowledge, fulfilled requirements and more.

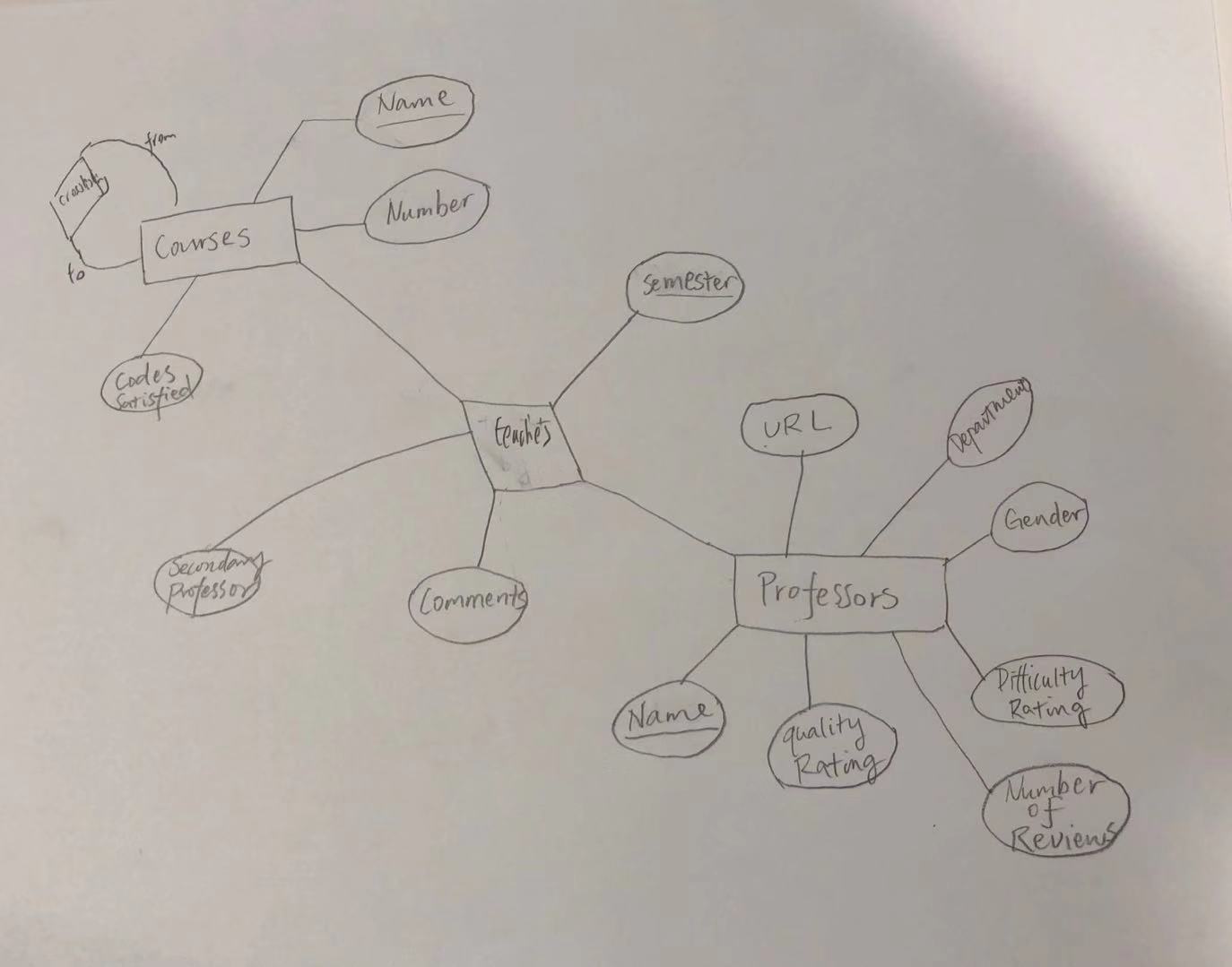
The brief overview/description of the course website for each class is very broad and general. It would be more helpful if its critical aspects such as levels of difficulty, average grades and student comments could be taken into evaluation. Our database system provides students the opportunity to see what their warm-hearted peers have to say about the list of courses they are about to dive into.

Plan for getting the data to populate your database:

We plan to get the data of our database from RateMyProfessor and Duke Academic Database. Some web-crawling work can be done to get the data from the RateMyProfessor website. The Duke Academic Database is already in XML format.

A list of assumptions:

1. We assume there are at most two professors teaching the same course during the same year. That is the reason why we included a secondary professor for our relationship table.
2. We assume the names of courses(ex. Introduction to Database) are unique keys for entity “Courses”.
3. We assume the names of professors are unique keys for entity “Professors”.
4. We assume courses are taught exactly once each semester.

The E/R diagram for our database design: 

The list of database tables with keys declared:

course(name, number, codes\_satisfied)

professor(name, quality rating, difficulty rating, gender, URL, department, number of reviews)

teaches(primary\_professor\_name, secondary\_professor\_name, course\_name, semester, comment)

crosslisted(course\_name,course\_name)

Web interface:

We want to design a website where users can find course and professor information that they are interested in. The website should allow users to view classes’ and professors’ reviews and other meaningful information. In addition, the class should allow comparison between entities such as classes or professors.

**Milestone 2 start point:**

7. Acquire the large “production” dataset: we have a large excel dataset that contains thousands of student rating, courses, and professors. For now, we plan to use the excel data to populate the dataset and continue to try to figure out how to grab data from ratemyprofessor.

8. Test the SQL statements you developed for Task 5 in the large database: The 3 populate sql files are able to populate our database, and most of the sql statements work, but we still need to write more sql commands in order to implement our design.

Summary of progress:

We completed all tasks from 1-8 but there are still more to polish and update for queries and database.

Because of the data we gathered is rather messy we need to further clean out the duplicates in the dataset and complete our web design. Other important things we need to do is finish our UI implementation, web development, and data population.