COSC 3318 Project Overview

General Description

The goal of the class project is to implement a database system application. The project includes the following activities spread over the entire semester:

- Identify an application area for which database systems may prove beneficial,
- Determine the functionalities of the database application (e.g., login, search, insert, delete, modify),
- Model the data stored in the database (Identify the entities, roles, relationships, constraints, etc.),
- Design, normalise, and perfect the relational database schema,
- Write the SQL commands to create the database, find appropriate data, and populate the database, and
- Finally and most importantly, write the software needed to embed the database system in the application.

The end result should be a functioning application that runs native and/or on the web and that uses your database to allow useful functionality.

A group of 3-5 students should do each project. You are free to choose your own project members; if you would like the instructor to assign you to a group, please send me an email at zhou@shsu.edu. There are 3 project assignments in total. You will get detailed instructions with each assignment. Each group should turn in a single solution to each assignment. Every member of the group will get the same grade.

Project Ideas

These ideas are just a sample. You are free to propose your own ideas. Realize that the ideas below are not complete descriptions. You need to work on them more and develop your project more concretely and in more detail. Do not get intimidated by the examples that are linked from this doc. These examples are meant to give you a feel for the application domain. It is up to you to narrowly define the scope of the application within the time frame of a semester-long project. Do not forget that you are supposed to have fun!

Project Ideas:

1. **Bibliography database:** Develop a system that will improve a research group's ability to track its publications and publications of interest to the group. Track information such as papers, authors, projects, conferences and journals. Readers should be able to view chronological listings, find papers by certain authors, group by projects, recover lists of papers based on keywords, etc. It should be easy for group members to add new papers, both written by the group and published by others in the literature. Examples of such systems include [CiteULike](#).

2. **Nobel Awards Database:** The goal is to model and populate information about the awards made in the various fields (Physics, Chemistry, Physiology or Medicine Literature, Peace and the Economic Sciences), the recipients, their countries, their year of birth etc. Your system should be able to answer questions such as "When was the first time an Asian won an award for the economic sciences?" (the answer to this particular question is 1998). The Nobel Foundation maintains such an [interface.](#) You could also work on variants of this idea such as the recipients of the [ACM awards](#) (unfortunately, there is not too much information online about this). Interesting queries then could be "Name people who have won at least two different awards" (the answer would include Knuth, Thompson, Ritchie, Engelbart etc.) Or the people "who were ACM Fellows before becoming Turing Award Winners" and so on.

3. **Books Database:** This domain is another popular one. Just look at [barnesandnoble.com](#) or [amazon.com](#) for excellent examples. You could model entities such as books, their authors, topics (which may be a complex hierarchy). You may also model various attributes of the authors and the institutions they belong to. You can support a service for buying and selling used books or books used in specific university courses. Your system can build a personal profile of people (and the books they like) and your database application could form the basis for a "recommender system", such as those supported by the commercial sites. The goal here is to "cluster" similar preferences together and the system can then make recommendations: "Since you liked *Harry Potter and the Sorcerer's Stone*, I recommend that you try *Harry Potter and the Chamber of Secrets*".

4. **Movies Database:** There are several excellent movie resources on the web, such as the [hollywood.com movies site](#) or the [Internet Movie Database](#). You could model entities such as movies, their actors, directors, genres, playing times, and reviews. There are several sources on the web from which you could get data to populate such a database. You can support various queries such as finding specific playing times, finding movies playing in Blacksburg directed by a given director. You can also support updates to the reviews section of the database (e.g., viewers giving their own opinions). Another functionality is to provide personal profiles of people (i.e., the movies they like) and then try to recommend movies to them based on profiles of viewers with similar tastes. You could also create a database of OSCAR or Golden Globe nominations and awards and answer queries such as "Find all the sitcoms that have been nominated three years in a row".

5. **Personal Photos database:** With the advent of cheap digital cameras, everybody has piles of digital photos. People need a way to organize, access, and show off their photos.

6. **Apartment Homes:** Our friendly neighbourhood web guide is here. This domain would require modeling apartments and their attributes, areas of town and their various characteristics (e.g., bus lines, crime rates, distance from various landmarks). You would provide an interface for offering apartments for rent, finding apartments based on various requirements ("gas heating + pets allowed + rent less than $500 + close to campus").

7. **Research Literature:** This domain involves modeling research publications. You need to identify the title of the publication, the forum it was published in, the authors, topics, keywords and related subtopic areas. This is a big business now (under the name of digital libraries). For example, the ACM digital library provides a beautiful searchable index (and retrievable repository, but that is beyond our scope) of nearly all of the publications of ACM. If you use this domain, then there are a lot of available resources for you to use. The ACM computing classification system provides a convenient hierarchical meta-index that you can use to organize your class hierarchy etc. If you are interested in a smaller domain, then the DBLP Bibliography Site provides a searchable facility for publications related to the database and programming communities. At the end of the day, you could identify papers written by a particular person at a particular place or ones in a narrowly defined area.

8. **Census Database:** Can you make a census data dissemination system for the Census Bureau? A census gathers data about people, business, geographic regions, etc. Different types of users need to gain different types of answers from the data. Homeowners want to know statistics about their region, such as crime rates. Business owners want to find holes in the competition. Government decision makers want to learn about demographic trends, and where to focus resources.

9. **Web Sites:** How do you think web search engines such as Google model their domain? You could think of them as a glorified database system where the basic entities modeled are web sites. You could then model the various properties of a web site: Topic, URL, domain name, other sites it links to, the background colour, etc. Retrieval could be for sites that have similar characteristics and properties.

10. **Others:** Of course, there are a whole host of other ideas such as bank accounts, student records, NBA data, election results, car rentals, auto insurance, consumer products and so on. Let your imagination run riot!