ECE 356 Project

Books and Library Data

This document is a description of details relevant to those doing the course project using the **Books and Library** datasets. It provides details about the dataset, as well as suggestions pertaining to the client application, the entity-relationship design, and the data-mining exercise.

Data Source

There are several sources of data for those doing a project in the Books and Libraries domain:

`https://www.kaggle.com/bahramjannesarr/goodreads-book-datasets-10m`
`https://www.kaggle.com/seattle-public-library/seattle-library-checkout-records`

You do not need to use all of the data in each dataset (*e.g.*, the library dataset contains several CSVs where each CSV is checkout records for a different year; selecting a single year is fine). However, it is extremely doubtful that you could do a good project if you only used one of the datasets as there are not that many distinct attributes in each set: Goodreads has only 23 distinct attributes, while the Seattle Library has 26. Neither has much complexity, but combined you can make a high-quality project.

All CSV files from the above sources have been made available on marmoset04 in "`/var/lib/mysql-files/20-Books/`" using exactly the names as on the Kaggle site.

In all instances, you should look at any of the relevant datasets on Kaggle to determine what the different attributes are within the CSV.

Client Application

As noted in the main project document, there will be little additional to add to the generic client-application requirements listed there. If you want a sense of what a client application for the books and libraries domain should do, you are advised to think about the potential users for such a database. The most likely users would be libraries and their customers. Such users would want to be able to borrow a book, keeping track of the records for that, so as to ensure that the book was return on or before its due date, look up whether or not a book was available, add new books to the database and remove those that were being disposed of, *etc.* It is highly likely that users would also want to check book reviews, so as to determine what books to acquire for the library and for customers to decide what books they wished to borrow. Adding or updating reviews is also something that would be desirable. It is not expected that you implement all such features, but, per the main project document, you should determine the desirable features in a client application and then select an appropriate subset to implement.

Entity-Relationship Design

Per the main project document, you will need to determine an appropriate ER design for your dataset. Prior projects done using this data have identified at least ten appropriate entity sets, as well as associated relationship sets connecting various of those entity sets together. If you have difficulty in thinking about different relevant entity sets for this domain you should consult with your designated instruction-team member.

Data-Mining Investigation

For book and library data there are numerous possible data mining-exercises that are worth considering. One of the most useful ones is expressed in the question, "Given a past history of book borrowing, possibly taking into account the length of time for which the book was borrowed, what would be good books to recommend the user consider borrowing?" Other questions of relevance include, "What factors (attributes) determine the popularity of a book?" and "Can reviews help in determining suitable book recommendations?"

If you have difficulty thinking about an appropriate data-mining exercise, you should consult with your designated instruction-team member. If you think you have a good idea for a data-mining exercise, it is probably worthwhile checking with your designated instruction-team member to confirm that it is of appropriate scope.