

CS 5300 Programming Project 1

Name: _____
Name: _____
Name: _____
Name: _____
Name: _____

The Antiquarian book collector who approached you earlier about creating a database for their collection of rare books is interested in progressing with your database design for her collection of rare books. She will provide to you the Excel spreadsheet she has been using for her inventory.

After examining this, it has occurred to her that some of the data she thought was recorded, is not in the Excel spreadsheet in a clear easy to extract fashion. For example, she does not have a column in the inventory for the language of every book. If you could try to get that information from the existing data, that would be great. Regardless, she has decided that after you get the data into the database she is going to have to go through her books and work on filling in the missing data. You kind of get the idea that, as a book collector, she might be thrilled to have this opportunity to go through her collection again. When you have loaded the data, she would like audits, or reports, to allow her to locate the books that are missing data.

Take your design and update it with your recent understanding of Normal Forms to create a design that is at least in BCNF and create the database (40 pts).

Document the design decision or changes from your original design (20 pts).

Take the Excel sheet, inventory.xlsx, that she has been using as her inventory and load her existing book inventory into the database (40 pts).

Document the problems you ran into and how you addressed them in a short description, or report, that you will email to the book collector (20 pts). The data is very messy and you will run into problems that you need to correct. How you choose to correct them is up to you, but be sure to document how you did it.

Create 4 reports: A report that lists all the Authors and the number of books they have (10 pts), a report that lists all the publishing companies and the number of books they have (10 pts), a list of books newer than 2000, which she is considering removing from her collection (10 pts), and a report that lists all the books that have missing data (20 pts). It would be great if these reports are in either Excel or comma separated value file format so they can be used as a checklist.

CS 5300 Programming Project 1

Notes:

You are most likely going to need to research another way to connect to the database to make loading the data easier for you. Common methods will require an Oracle client installed, and/or a language specific library. If you want to use Java, you would only need to use the JDBC library, but no Oracle client. If you want to use VB or C#, you would need the Oracle client and configure an ODBC connection to the database. Some tools and languages, python and perl, will use the Oracle client through native libraries (OCI), and language specific libraries such as (cx_Oracle, or DBI). I think C will use the OCI libraries from the Oracle client directly.

To deal with the messy data you might want to preprocess the data in your language of choice, or you might want to load the entire inventory into intermediate tables and massage the data into a workable format before loading them into your final tables.

Authors can be multivalued, so be sure to treat them as such. This means taking the author string for multi-author books and breaking it up into a list of individual authors.

You may use the Oracle database that is available to you or you can use a database of your own as long as you can provide a way for me to tell that it is real and not vaporware.