Instructions

Overview

This assignment requires you to take a dataset of your choice, to analyse and model it, and then to implement it as a database with a simple associated web application. The assignment consists of four stages:

Find a dataset and critique it, identifying some questions you would like to ask it;

Build an E/R model for the data structure and adjust it for relational implementation;

Build a MySQL database for the structure and put some or all of the dataset into it as instance data;

Construct a simple web application in node to illustrate the data and implement asking some of the questions you identified in the first step. You will need to install packages as you did in your previous work – no package files have been provided for you.

You will submit a report in .pdf format, and the link to your lab environment with the MySQL database and the web app in it (There should be a big button for a shareable link on page that launches the lab).

Although there is no page limit for your report, we recommend that you aim for less than 20 pages (excluding any code, data and screenshots). In practice, reports far shorter than that often get top marks.

Don't forget to include both the link to your lab and your pdf – we will not be able to mark work that we can't access.

More information on each of the four stages is given in their sections below.

Stage 1. Find and critique a dataset

Choose a source of open data. This can be one that you've used earlier in this module, or it can be a new source. You may also combine two simpler datasets, provided there is benefit in linking them. The only restrictions are: the data must be open; a normalised, relational model for the data can't have been published already.

Assess the dataset in your report. You should use the criteria used in discussion 1.104 (Quality, level of detail, documentation, interrelation, use, discoverability) and assess the terms of the use (as in discussion 1.206)

Quality: Is the data reliable? On what basis do you make that judgement?

Detail: How much detail is there? Is the information helpful?

Documentation: How clear is it what the data means? Where was the documentation? Was it easy to find?

Interrelation: Would it be useful to connect to other data sets? Which? How easy would that be?

Use: What could you use it for? What questions would you like to ask the dataset but can't? What's missing?

Discoverability: How easy was it to find open data in your chosen domain? Where did you go? Were there many alternatives?

Explain your interest in your report. Why is this an interesting dataset? Give some questions you would like to ask of the dataset that a database application could help with.

Stage 2. Model your data

Draw a complete E/R model of the data. If you are only implementing a subset of the data structure, justify your decision in the report. All diagrams must be in the report pdf.

Add cardinality to your E/R diagram. If there are any structures that are not compatible with the Relational model, draw a second diagram showing the modified structure.

List database tables and fields. Evaluate the tables against the normal forms and adjust to ensure that your database will be at least in 3NF. Include an evaluation of which normal forms your database is in. If it isn't in higher forms, such as BCNF or 4NF, justify your decision not to normalise further.

Stage 3. Create the database

Build the database structure in MySQL in the lab environment. Record all CREATE commands used in your report.

Enter instance data. This can be a usable sample of the dataset or all the dataset. Detail how you added the data in your report.

Reflect on how well the database reflects the data. In your report note one or two points of elements that do or don't work well.

List SQL commands that answer questions identified in Stage 1/Step 3. If any can't be answered, explain why.

Stage 4. Create a simple web application

Write a node.js module to present a web application that queries the database. This can be quite simple, but should address some of the motivation and questions identified in Stage 1/Step 3. The user account used to connect should have appropriate privileges

Take screenshots of the main screens from your web application and include them in your report.

Review Criteria

Each stage has equal weight in the marking scheme.

The evaluation criteria for each stage are given below. Each criterion is worth 5% of the total marks, and you should expect to meet at least two criteria fully in each stage to pass.

Stage 1. Find and critique a dataset

Dataset is appropriate

Dataset assessment criteria are all addressed

Dataset assessment is perceptive

Interest in dataset is well justified

Questions are identified and justify a database approach (simply looking up values in a single table may not justify the complexity of a relational database)

Stage 2. Model your data

E/R diagram identifies all fields and entities to be modelled

E/R diagram is legal

E/R model is sensible and any issues with the relational model are resolved

The normalisation analysis is accurate

The database is appropriately normalised

When you draw your E/R diagram, follow the rules from this module and its set reading. Lots of other things get called E/R diagrams (you may have met other ones for other modules). Only diagrams that match the ones in this module will get full credit.

Stage 3. Create the database

The implemented database reflects the E/R modelling

The database is sensibly implemented (including column types)

Keys are correctly identified

Reflection is good

Queries are correct and accurately reflect the identified questions

Stage 4. Create a simple web application

The application runs

Connections and queries are appropriately handled

The data is presented appropriately

Generated HTML is valid

Motivation and (some) queries are addressed