CMPT 308L-111: Lab 1

Professor Labouseur

Graham Burek

January 26, 2016

Data vs. Information

Select a database in use today (real or imagined) and identify the elements of data stored therein and describe how the database organizes the data into information. Give contrasting examples of data and information that illustrate the meaninglessness of data without context and organization. Talk about the value the information provides once the component data is given context.

Suppose there is a database for a particular McDonald's franchise. Inside this database are pieces of data ("Fries", "Large", 1.49, for example). By itself, this data is not useful. While this data in the database is not random (it has a meaningful ordering), it is not useful because it lacks context. Translated into the terms of an actual restaurant, having raw data without context is analogous to a business giving customers random numbers about the prices of the products they are selling. This data pertains to the business, but it is incomprehensible in its current form. Information, on the other hand, is data given context. In the terms of a database, information is data given some sort of structure and relationship. In a relational database design, the data is organized into tables, and relationships are made by linking tables together by their keys. In this form, data becomes useful. Using the restaurant analogy, giving a customer information would be akin to giving them a menu with the names, prices, and nutritional information of the food the restaurant serves.

Data Models

Briefly describe the hierarchical and network pre-relational data models. Explain their shortcomings in relation to the relational model. Considering this, what do you think of XML as a model for data storage?

Wanting a more abstract, platform independent way to represent data on a computer system than the file system provided, the hierarchical data model was created. The hierarchical model is a data model that expresses relationships between its entities by using a tree model. An entity can either be the root of the tree, or have one parent (higher level relationship) and any amount of children (lower level relationships). The tree can be queried to find out how data is related. Unfortunately, this design forces the duplication of data (if the data belongs to 2 different nodes on the tree), possibly creating data inconsistency. The network model remedies this problem by allowing nodes to share the same children, and have more than one parent node. Some problems with the networking model still remain, however. Network models are meant only to represent one kind of entity at a given depth. This means that if two entities share a common parent but are unrelated, the network model must be used in an unintended way in order to represent this relationship. The relational model, on the other hand, does not have any of these problems. By giving every entity a separate table, and by representing the relationships between the entities in tables, the relational model fully accommodates all types of relationships. Considering the setbacks of the hierarchical model, using XML as a model for data storage only makes sense if the data to be stored is a group of similar entities with unique, non-overlapping children.