Kramer Johnson

CPSC 5021 02 Fall 2020

9/18/20

1. **A Summary of *The Open Society***

Over the past few decades, data has come to take a large role in all our lives. We have all noticed this and a lot of attention is being paid to how it is collected, how it is used, and how much of it is disclosed. When most people think about issues around data these days, they tend to think about businesses in the private sector. *The Open Society* is an article about how data is being handled in the public sector.

From crime data, to pollution data, to legal data, the authors of this article argue that open policies toward data held by the government is for the greater good. They provide examples such as how open access to crime data has lead to websites such as "San Francisco Crimespotting" that let users see historical crime data layered on top of maps and FixMyStreet.com in Britain that lets citizens bring attention to local issues such as graffiti, broken sidewalks, or street lighting (http://www.fixmystreet.com).

The main point of this article is emphasized by the quote from Vivek Kundra, that providing access to data "creates a culture of accountability". When governments and companies open their databases to the public, it prevents abuses of power and law. Such as the example of American businesses reducing emissions by 40% after being required to disclose their data on pollutants. Whether the data is released raw, and citizens process it into information, or the processed information itself is made public, there is little downside to public access to public data.

2. 1. Accounting database of invoices at work
   2. Mountainproject.com database of rock-climbing routes
   3. My gmail account's database of emails
   4. Wta.org's database of Washington hikes
   5. A gas station's database of customer purchases
   6. Reddit's database of users
   7. REI's database of current products
   8. My stockbroker's database of current and historical stock prices
   9. ESPN's database of fantasy football statistics
   10. Discogs.com's database of vinyl record releases

3. Group Members: Kramer Johnson, Anthony Pinza, Jun Liao

   Group Leader: Kramer Johnson

Project: A social travel website

Explanation: Our plan is to have a database for a social travel website. This site would not only have a database of popular travel locations, but it would keep track of it's members, where they have been, how they rated those places, what places are on their "want to see list", and have the ability to connect those members with others who liked similar places and want to go to similar places. The problem we are trying to solve is that people enjoy travelling with others but may not know anyone to do so with and we can provide travelers with destinations they may enjoy but would not have otherwise known about.

Review Questions

1.

1. **Data**: Raw facts, or facts that have not yet been processed to reveal their meaning to the end user.
   Field: A character or group of characters that has a specific meaning. A field is used to define and store data.
   Record: A logically connected set of one or more fields that describes a person, place, or thing.
   File: A collection of related records.

4. A database management system is a collection of programs that manages the database structure and controls access to the data stored in a database. The main functions of a DBMS are data dictionary management, data storage management, data transformation and presentation, multiuser access control, backup and recovery management, data integrity management, database access languages and APIs, and database communication interfaces.

7. The role of a DBMS is to manage database structure and control access to the data in the database. Specifically, it stores metadata about the database and enforces the rules contained in the metadata. Some advantages of a DBMS are improved data sharing, improved data security, better data integration, minimized data inconsistency, improved data access, improved decision making, and increased end user productivity. The disadvantages of a DBMS are increased costs, management complexity, maintaining currency, vendor dependence, and frequent upgrade/replacement cycles.

10. Meta data is data about data. It is data about data characteristics and relationships.

13. Some examples of unstructured data could be emails, audio, or books. Structured data can be names, zip codes, cities, or prices that have been organized into a table. Unstructured data has not been organized (structured) in any way to make it useful for processing into information. Structured data, on the other hand, is the result of formatting unstructured data for storage or processing into information. Unstructured data is more prevalent in a typical business environment due to the large volumes of emails, invoices, and voicemails that are exchanged within the organization and with outside parties.

Problems

2. If you wanted to sort by city you would need to parse the string that is contained in the MANAGER_ADDRESS column. This would involve a procedure to detect what part of the string is a city, then storing that value for each record, and then sorting the records by those values. To fix this issue, it would be best to separate cities of residence into a separate column titled CITY then you could easily sort by that column.

5. In this table we can see that project names, project numbers, job codes, and job charge rates are all repeated. It is redundant to have a project number and a project name in the same table, to have employee numbers and names, and to have job codes and the job charge rates all in the same table. Also, employees are repeated on separate lines creating duplicate entries that could lead to serious data inconsistency problems.

6. For the EMP_NAME and EMP_PHONE contents I would separate the first and last names of each employee into separate fields. I would also have only one entry for each employee (preferably in a separate EMPLOYEE file) so that their names and phone numbers are not duplicated in different records.

7. The data sources in the file are projects, employees, and jobs.

8. To eliminate data redundancy, I would create a PROJECT file, an EMPLOYEE file, and a JOB file. This way each PROJECT, EMPLOYEE, and JOB would only be listed in one location and could be linked to the associated PROJECT, EMPLOYEE, or JOB via foreign keys to show their relationships.