

Intro to Databases

csci 350

GIS data is stored in a database. Part of this class we overview the fundamental issues in databases, namely data models and database system architecture. We'll look in more details at the relational model.

Here is a set of readings. You can find them linked on the class website (they are available on the web,).

1. E. Codd. *A relational data model for large shared data banks..* In *Communications of the ACM, 1970*.
2. M. Stonebraker and J. Hellerstein. *What goes around comes around..* In *Readings in Database Systems, MIT Press*.
3. MYSQL and Open Source [search the web].
4. supplementary: M. Stonebraker and J. Hellerstein. *Anatomy of a Database System*.

As you read these papers, think about and be prepared to answer and discuss the following questions next time in class:

1. What is the notion of *data independence*? Why is it important?
2. Why is the relational model superior to the ones before?
3. What are the key ideas behind the relational model? Why/how it improves previous models?
4. What is a normal form (in a relational model)? Why is it important that a database be stored in normal form?

5. What does Codd mean by *derivability* of a relation?
 6. What ways does Codd propose for detecting inconsistencies in a database?
- IMS What is the IMS model? What are the options for storing roots in IMS (pros and cons for each)? Does IMS have data independence?
- CODASYL How does CODASYL organize data? Give an example of data that is unnatural to model with CODASYL. Compare Codasyl and IMS. Why does Codasyl resemble “navigation in hyperspace”?
- R What was CODD’s motivation and goal for introducing a (new) model? What decided at the time the “win” of relational DBMS over Coodasyl/IMS?
- R++ What was the tradeoff with the extensions to the relational model? What do you think of Lesson 12?
7. What does it mean “impedance mismatch” of a relational database?
 8. problems with OODB.
 9. What was the motivation of the OR model?
 10. Reflect on the data classification in the paper: rigidly structured, rigidly structured with free fields, semi-structured, unstructured. Give examples of each.
 11. XML is a model for unstructured data. Name some problems of XML. What do the authors think of XML? Future?
 12. Why is the article called “What goes around comes around”?
 13. What is MySQL? Why is it important? What does Open Source mean? What is the difference between free software and open source software? Who is Richard Stallman?