

## CS 107 Fall 2006

### Lectures 9 and 10: Database Concepts

**A database is . . .**

- A collection of records
- that can be searched using a query language such as SQL (pronounced sequel).

A *flat database* is much like a spreadsheet: it has only one table. A *relational database* can contain multiple tables of related data. The database engine is used to correlate information from them. Professional databases support concurrent access by multiple users, and use locking protocols to ensure that two users cannot interact in a way that damages the data.

- One *database* can contain several tables storing related data.
- A *table* is a list of records
- A *record* is all the data about one person or transaction or thing.
- Each record can have many *fields* such as name, age, price, etc.
- Each field is one or more *characters*. Examples of fields: 3.25, James, CT

**Similarities.** In some ways, a flat database is like a spreadsheet:

- They both store heterogeneous records, that is, a record can contain strings, numbers, codes, etc.
- Both can be searched.
- You can produce management reports from the data stored in either one.

**But a database is not just a fancy spreadsheet.** Some of the important differences are:

- The fields in a database record are grouped permanently together, while the fields in a spreadsheet record are just independent data.
- A database lets you edit only one record at a time. You cannot accidentally modify one database record while intending to modify another.
- A spreadsheet will let you enter anything into any cell, while a database forces you to enter the right kind of data into every cell. You cannot put a string into the first field of one record, and a number into the first field of the next record.
- In a database, when you search, you search a specific column, not all the data everything. When you search a spreadsheet, you might find the search string in any column.
- A database lets you specify the kind of search you want: search for exact equality, less than, greater than, or strings that are “like” the search string but not identical to it.
- In a database, when you search, you get all of the matches instead of one at a time.
- You often sort a spreadsheet, while a database is stored in whatever order the engine has chosen. You don’t see that order. However, you can sort the results of a query.

## A database management system (DBMS) has two parts.

- An *engine*. This is a large and complex computer program that stores the data and does the retrieval. The Microsoft engine is called **Jet**.
- An *interface* or "front end". Microsoft's supplies a graphic user interface called Access as part of Microsoft Office. Behind the interface is the Jet engine.

The word "database" is sometimes used to refer to one or both of these parts, or to a collection of records that is stored in the system.

## Responsibilities of the interface.

- To let you define your tables (your record formats).
- To let you enter the data.
- To let you fix errors.
- To let you form queries for the engine and report the results of those queries
- To support reports that are used for management decisions. The interface must get the data for the reports by querying the engine.

We will look at a GUI interface that is very much like Access and examine how it supports each of these responsibilities. The software I am using in class is called NeoOffice; it is a free Mac version of Open Office, a free software package for Windows and Linux systems.

**Defining a table.** Before you can create a database you need a clear idea of the kind of data you will store in it. When you create a table, it is easy to add as many data fields as needed. After you enter the data, changing the structure of a table is a pain, at best, since all existing records must be modified.

When you create a table, you start by naming the fields and specifying the type of each field. Databases usually support several types of data, including the ones listed below and others.

strings	first name, last name, middle initial	Mary, J
numeric strings	zip code, academic major code	06516, 357
numbers	grade-point average, age	3.54, 21
dates	October 13, 2006	
codes	gender, state	M, F, CT, NY, NJ, etc
boolean	used for T/F situations	yes or no
auto number	used to create unique ID numbers	0, 1, 2 ...

**Why so many types?** Strings are for representing alphabetic information and numbers on which you do not intend to do arithmetic. Numeric strings are used for things like telephone numbers and zip codes, which might be use for comparisons but never for arithmetic. Codes are different from strings because they come from finite sets of allowable symbols. Boolean values form a code with only two values, but they are so common that many databases support a special type for them. Dates can be represented as strings or as numbers, but the require a special and complex kind of error checking (validation).