

More about Databases

Computer Literacy 1 Lecture 18
30/10/2008



Topics

- Terminology
- What's a database for again?
- Relational Database Model
- Samples
- Database Integrity
- Data mining



More Taxonomy

- According to more orthodox database terminology a *file* is a collection of **recordings**
- A record is the information relating to a single item of the relevant type, e.g. person, product, event ...
- In an address book database, a record contains information about **1** person, a CD catalogue would have **1** record per CD



And more Taxonomy

- Each discrete chunk of information in a record is called a **field**
- The type of information a field can hold is determined by its **field type** or **data type**
- For example the author in a library database would be a **text field**, the field for specifying the number of copies would be a **numeric field**



Other fields



- A date of purchase field might be a **date field** that only allows dates to be entered
- Many database programs allow fields to contain graphics, photographs, sounds, video clips
- **Computed fields** contain formulas similar to spreadsheet formulas: They display values calculated from other *numeric fields*

Layers



- Recall layers of abstraction architecture:
 - Physical, logical, external

The abstract logical layer is the heart of the database since this is where the **data structures** live

What is a database for



- Ultimately, the point of a database is to model part of the **real world**
- So we want the data structures to represent the corresponding area/realm of **facts**
- The realm of facts is a complex place: **objects** have many different *attributes* and bear many different *relationships* with each other
- The model should accurately express these attributes and relationships

Different Types



- There are various methods for structuring the data to represent the facts, e.g. flat files, hierarchies, networks, object oriented models
- Our focus will be on **relational model** - currently most wide spread

Different Types



- The DBMS transforms data in the physical data files into a common structure in the logical layer (*schema*)
- Different data structures are generated by different DBMSs

What's a database again?



- Structure - Schema
- Data
- Within the structure we can define classes of items, their properties and the relationship between them
- Data is then entered into the database to conform the schema
- You can have a schema with no data

Database vs. DBMS



- A database is a schema **plus** data
- The program that lets you work with the data is a Data Base Management System (DBMS)
- MS Access are DBMSs, not databases

Relational Database Model



- Developed by E.F. Codd, 1970
- Only one basic structure in RDB: the **relation**
- The relation is a table of objects of the same type
- With properties:
 - Every relation must have a unique name
 - Every column in it must have a unique name
 - Duplicate rows are not allowed
 - Enforced by a requiring a unique primary key
 - Each cell must contain a single value

Example - Structure



- Schema: Media catalogue
- Instance: My Music Collection
- Classes: Artists, Work
- Relationship: Artists record work etc...
- Attributes:
 - Artists have names, category, nationality, ...
 - Works have titles, dates, media, ...

Example Data



- Artists:
 - Eels
 - Kings of Leon
- Works
 - Beautiful Freak
 - Only by the night
- Relationships
 - The Eels **recorded** Beautiful Freak
- Attributes
 - Beautiful Freak **was recorded** in 1996 **on CD**

Database Integrity



- Consider:
 - Only by the night Kings of Leon
 - Only by the night King of Leon
 - Aha Shake Heartbreak King of Leon
- Variations produce inauthentic duplication and lack of association (a search on Kings of Leon would miss Aha Shake Heartbreak)
- This database is not an accurate reflection of the world it represents: it lacks **integrity**

Relational Database



- Database comprises a **set of tables**, where all the data on one topic is stored in a table
- A record is the data for a for a single item in a table, e.g. the details relating to one artist
- A field is a piece of data within a record, e.g. album title
- A relationship links detail in one table with detail in another, e.g. through a ArtistID number

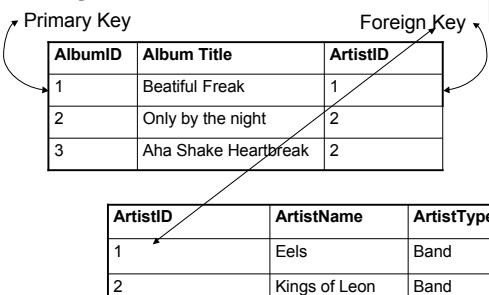
Tables

- Each **record** in a table is presented in a *row*
- Each **field** in a record is in a *column*
- Each *field* has a **field name** at the top of the column
- The type of information a field can hold is determined by its **field type**, e.g. *text* field, *numeric* field, etc...

Keys

- Most records will have a **unique identifier** - a field that holds different information in every record
- The field that must be unique in each record is the **Primary Key** - you *can't enter duplicate information* in a primary key field
- A **Foreign Key** refers to a Primary Key in another table

Sample Relations



Join Albums with Artists

- (Natural) join of Album and Artists

AlbumID	Title	ArtistID	ArtistName	ArtistType
1	Beautiful Freak	1	Eels	Band
2	Only by the Night	2	Kings of Leon	Band
3	Aha Shake Heartbreak	2	Kings of Leon	Band

A Few Relational Operators



- Restrict
 - Returns selected rows according to criteria
- Project
 - Returns selected columns according to criteria
- Join
 - Returns one relation from 2 (more columns)
- Union
 - Merge 2 compatible relations (more rows)
- Intersection
 - Returns rows common to 2 relations

The External Layer



- Uses a query language, e.g. SQL (also handles schema definition) based on relational operators
- Input via direct entry into a table or via a form
- Query as text enquiry or via GUI
- Output as response to a query
 - A user view

Data Mining



- Discovery of hidden/unexpected patterns of data
- A.K.A. Knowledge Discovery
- Involves large volumes of data
- Process of sorting through large amount of data and picking out relevant information
- Useful in making organisational decisions

Data Mining Techniques



- Predictive modelling
 - Test model against real data
 - Or train your model on a sample data set
- Segmentation
 - Look for clustering in data
 - Buying patterns

Key points

- Terms: Schema, relations, DBMS
- Relational Model and how it is applied
- Database integrity
- Selected operators and what they do
- Data mining

