



University of  
**Southampton**

# Data Types

COMP3211 Advanced Databases

Dr Nicholas Gibbins - [nmg@ecs.soton.ac.uk](mailto:nmg@ecs.soton.ac.uk)

# Overview

- Data types and operations
- Temporal data
- Spatial data
- Multimedia data

# Data Types and Operations

# Data Types

- Numeric
- Character
- Temporal
- Spatial
- Image
- Text
- Audio and Video

# Operations on Data

- Comparison
- Arithmetic
- Fuzzy searches
- Retrieve all documents that contain a given word
- Find a picture that contains blue sky

# Which operations are meaningful?

Can you add two weights together?

- $2\text{kg} + 2\text{kg} = ?$

Can you multiply two weights?

- $2\text{kg} * 2\text{kg} = ?$

Can you add a weight to a quantity?

- $13 + 2\text{kg} = ?$

Can you multiply a weight by a quantity?

- $13 * 2 \text{ kg} = ?$

# Which operations are meaningful?

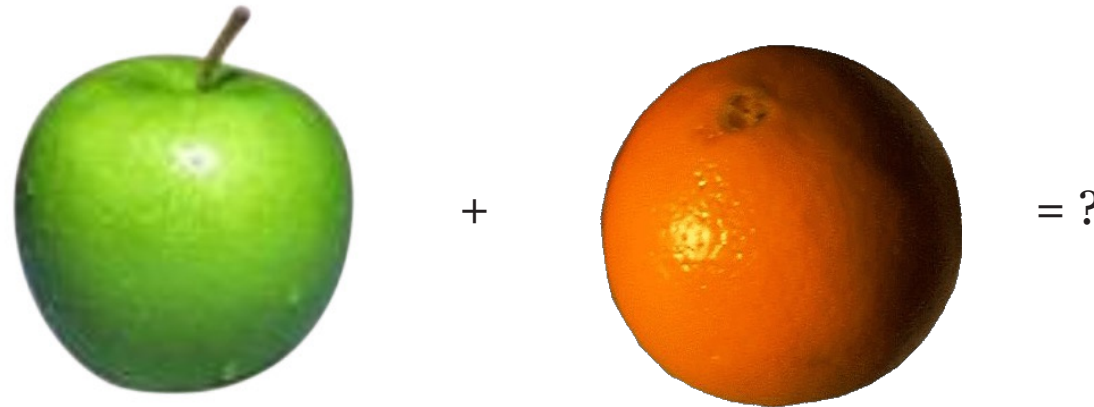
Can you compare two images?





# Which operations are meaningful?

Can you add two images?



# Further Questions

Is the data ordered in any sense?

- Total order vs. partial order

Does the order actually have any meaning, or is it just a convenience?

# Temporal Data

# Temporal Data

The dimension of time is needed to answer such questions as:

- What was the average price of product X during 1995?
- In which month did we sell the most copies of video Y?
- What was the treatment history for patient Z?

# Characteristics of Time

## Time structure

- Linear
- Branching time, possible futures
- Directed acyclic graph
- Periodic/cyclic

## Boundedness of time

- Unbounded
- Time origin exists
- Bounded at both ends

# Time Density: Discrete

Timeline is isomorphic to the integers

- Integers have a total order

Timeline is composed of fixed periods, termed *chronons*

Between each pair of chronons is a finite number of other chronons

# Time Density: Dense

Timeline is isomorphic to the rational numbers

- Rational numbers have a partial order

Between each pair of chronons is an infinite number of other chronons

# Time Density: Continuous

Timeline is isomorphic to the real numbers

- Real numbers have a total order

Between each pair of chronons is an infinite number of other chronons



# Characteristics of Time

Granularity is important

- Event A occurs at 11.00am
- Event B occurs at 3.00pm the same day
- Does event A precede event B?
- The answer is different if
  - Granularity is one day
  - Granularity is one minute

There is also a distinction between sequence and time

# Storing Times in a Database

Various times may be associated with an event that appears in a database

We may wish to record

- The Valid Time of a fact – when the fact is true in reality
- The Transaction Time of a fact – when the fact is current in the database, and can be retrieved
- Both of these (*bitemporal*)

# SQL Extensions

TSQL includes:

- A WHEN clause (see next slide)
- Retrieval of timestamps
- Retrieval of temporally ordered information
- Using the TIME-SLICE clause to specify a time domain
- Using the GROUP BY clause for modified aggregate functions

# TSQL WHEN Clause

Format of the SELECT ... WHEN statement

- SELECT { select-list }  
FROM { list of relations }  
WHERE { where-clause }  
WHEN { temporal clause }

Temporal comparison operators include:

- BEFORE/AFTER, FOLLOWS/PRECEDES  
DURING, EQUIVALENT, ADJACENT, OVERLAPS
- (compare with Allen's Interval Calculus)

# Spatial Data

# Spatial Data

Data Types include:

- Points
- Regions
  - Boxes
  - Quadrangles
  - Polynomial surfaces
- Vectors

Operations include:

- Length
- Intersect
- Containment
- Overlap
- Centre

# Spatial Data Applications

Computer Aided Design (CAD)

Computer generated graphics

Geographic Information Systems (GIS)

For these systems, the properties of interest would include:

- Connectivity
- Adjacency
- Order
- Metric relations

# Spatial Data Characteristics

In systems dealing with space:

- Data objects may be highly complex
- Data volumes may be very large
- Data may be held in real time
- Performance is not easy to achieve
- Access is likely to be through specialised graphical front ends; operator skills are key
- Query processing will (probably) not be performed using SQL



# Multimedia Data

# Textual Data

Text data may be

- Already in machine-readable form, from a word-processor, spreadsheet or other source
- Read using OCR techniques

Text data is essentially unstructured, and an index of some kind needs to be built

- By a human operator
- Automatically by building an inverted list of every significant word in the database

# Textual Data

Markup languages do give some structure to a document

- HTML is a markup language for the Web

XML (and its predecessor SGML) allows a programmer to create portable documents that contain structured data

- Can also create new markup languages

Character Large Objects (CLOBs) are now commonly supported by vendors

- Able to store and handle text documents in addition to standard data
- Provision of text search and retrieval facilities

# Text and Documents

Much data is stored in the form of text

It would be very useful to be able to ask queries such as:

- Find all the legal documents concerning client 'Jones'
- Find all the suspects with false teeth who have been interviewed
- Find all the articles on 'databases'

# Image Data

Examples of still images include:

- X-Rays
- Maps
- Photographs

These are all classified as binary large objects (BLOBs)

- No attached semantics

# Image Databases

An image database needs to provide support for:

- Image analysis and pattern recognition
- Image structuring and understanding
- Spatial reasoning and image information retrieval

Mainstream DB vendors now adding

- Support for BLOBs
- Access using QBIC (Query by Image Content)

# Audio Data

## Digitised sound

- Stored in various formats, such as WAV or MP3
- Consumes large amounts of storage
- Compression techniques normally used

## MIDI (Musical Instrument Digital Interface)

- More compact than digitised audio
- Consists of a sequence of instructions:  
Note\_On, Note\_Off, Increase\_Volume
- Interpreted by a synthesiser

# Video Data

One of the most space hungry formats of all

- Images stored as a sequence of frames
- Each frame can consume over a megabyte
- Frames typically played back at 24-30 fps

To integrate video and audio, interleaved file structures incorporate times sequencing of audio/video playback

- Microsoft AVI
- Apple Quicktime



Next Lecture: DBMS Architecture