

F27WD: Web Design & Databases

Databases Lecture 1:

Data - what is it good for?

Fiona McNeill

18th February 2019

My learning objectives*

- Understand that data and data management is *important, relevant and exciting*
- Understand how data is used and why you should care
- Be able to do your own (simple) things with data.

* check the Course Descriptor for the technical learning outcomes

An interactive course

- We will be using Google forms to allow interaction. The link to the lecture's form will be available on Vision, and will also be available in the lecture.

This week's link:

<https://goo.gl/forms/Y4J2B91MRa7M9cpp1>

Q1: Can you use this form properly?

1. Yes, here I am!
2. I'm trying to make it work but it's confusing
3. What was the question again?

An interactive course

- I will also pick on people and ask questions!
 - *I don't know* is an ok answer
 - Not all answers are right, but most have something interesting in them.

Now the learning begins ...

Qu1: I think that data is

1. Fascinating and exciting
2. Pretty dull
3. Not really sure but prepared to be convinced

What is data?

What is data?

Broadly:

A datum is a piece of information; something that is known. Data is a collection of these.

In computing:

Information that is machine processable, usually stored in binary format.

Data - information - knowledge

Data - information - knowledge

One way of looking at this:

Data: student 473 ordered item 11736 at 12:34:03

Data - information - knowledge

One way of looking at this:

Data: student 473 ordered item 11736 at 12:34:03

Information: Dave had a ham sandwich for lunch

Data - information - knowledge

One way of looking at this:

Data: *student 473 ordered item 11736 at 12:34:03*

Information: *Dave had a ham sandwich for lunch*

Knowledge: *Dave is not a vegetarian*

Data - information - knowledge

One way of looking at this:

Data: *student 473 ordered item 11736 at 12:34:03*

Information: *Dave had a ham sandwich for lunch*

Knowledge: *Dave is not a vegetarian*

But these words are used in lots of overlapping ways.

Types of data

- Raw data
- Processed data
- Experimental data
- Field data
- Etc

These may be overlapping concepts

The World of Big Data

90% of the data in the world has been created in the last two years¹

What does this mean for us?

¹<https://www.sciencedaily.com/releases/2013/05/130522085217.htm>

The World of Big Data

This affects almost every aspect of life, e.g.,:

- Medical advances - analyse every case in the world to get personalised diagnoses
- Personalised service - understand what your customers want and give it to them
- Politics - make people think what you want ...

The World of Big Data

- Politics - make people think what you want ...

In the 2012 presidential election, the Obama Campaign made use of voter models on a scale never before seen. They were able to identify specific voters who would make a difference in the election and target messages to those voters. I am not talking about something general like, “we need to appeal to soccer moms,” I am talking about true specifics like, “the Johnson family Maple Lane in Columbus, Ohio will vote for us if they know our stance on social security.”¹

- Nate Silver predicts elections better than other people because of how he uses data.

¹<http://insights.wired.com/profiles/blogs/5-ways-big-data-will-change-the-world#axzz4zjAZ91Kx>

Who has data about you?

Who has data about you?

- Google
- The Government (UK, Scotland, local, etc)
- Heriot-Watt University
- Your bank
- Your local pizza place
- Most places you've ever shopped
- A **lot** of other places!

Living in the World of Big Data

Do you want to be a **passive** participant or an **active** participant?

Managing data is **hard** - no one knows how to deal with data on this scale.

But there is a lot we do know and a lot we are learning. This course will introduce you to the basic ideas, from which everything else can grow.

Managing data: Databases

So we have all this data - what are we going to do with it?

Databases

We use databases to **store, manage and access** data

Databases

We use databases to **store, manage and access** data

A database is an *organised collection* of related data

Databases

We use databases to **store, manage and access** data

A database is an *organised collection* of related data

It is a *model* of some restricted but well-defined part of the world.

Databases

We use databases to **store, manage and access** data

A database is an *organised collection* of related data

It is a *model* of some restricted but well-defined part of the world.

It *represents facts* about this model.

Designing a database

If we have lots of data that we want to be able to use and access, we have to *structure* it.

The first thing to do is to decide what the structure will be like.

 **Let's watch a film ...**

What sort of film should we watch?

What sort of film should we watch?

We might categorise films by things like:

- Genre
- Stars
- Director
- Length
- Release date
- Rating
- Reviews

What sort of film should we watch?

Note: there are a more or less unlimited number of ways we could do this - e.g.:

- Number of cinemas shown in
- Person who made the tea on set
- Number of costume changes
- ...

Some of these are more *intuitive* or *common* than others

Developing the database: MyFilms table

A *table* is a simple way of storing data.

First of all, we need to decide what kinds of things we want to say about the data.

This is the **schema** of the database and it models the data.

Developing the database: MyFilms table

filmName	director	genre	rating	year

First, we develop the schema: *what kinds of things can we say about this data.*

There are lots of other ways we might have chosen to do this.

Developing the database: MyFilms table

filmName	director	genre	rating	year
Ghostbusters	Paul Feig	Comedy	12A	2016
Gone with the Wind	Victor Fleming et al	Drama	PG	1940
Star Wars Episode IV	George Lucas	Sci-Fi	U	1977
Die Hard	John McTiernan	Action	18	1988

Next, we populate the schema with the data itself.
Each line is an *instance* or a *row*.

How do we create this?

There are lots of different ways.

In this course, we will mainly look at:

Relational databases

Designed using Structured Query Language
(**SQL**)

Using **MySQL** as the Database Management
System (DBMS)

Next lecture ...

- We will start to build our own database!

Course logistics

- Lectures will be fairly short (40 mins ish) and interactive
- Labs are essential
- You need a MySQL account
 - You should have an email about this
 - If not, let me know asap!

Assessment

- You'll be getting further detail of this in a week or so
- You'll be working in teams to build a database
- It will be due in week 12

How to contact me

I work part-time (all day Monday; Tuesday, Wednesday, Thursday till 1430; not on Fridays)

Best: Office hour: Wednesdays 1215-1315 in EMG:52

Usually fine: Talk to me after lectures

A bit slower: Email me: f.mcneill@hw.ac.uk

I will aim to reply within 2 days