OCR Computer Science AS Level

1.3.1 Databases

Advanced Notes

Specification:

1.3.2 a)

- Relational Database
- Flat File
- Primary Keys, Foreign Keys, Secondary Keys
- Entity relationship modelling

1.3.2 b)

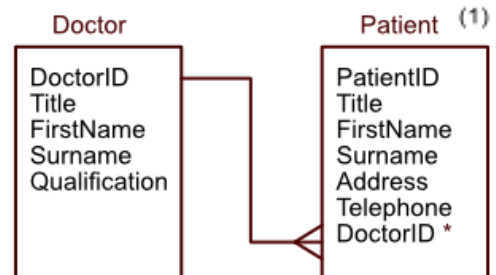- Methods of capturing, selecting, managing, and exchanging data

Relational Database

Relational Databases

An entity is an item of interest about which information is stored. A relational database is a database which recognises the differences between entities by creating different tables for each entity.

The diagram on the right is an entity relationship model which shows two entities: Doctor and Patient. DoctorID is the attribute linking the two tables together. Attributes are characteristics of the entity; these are categories about which data is collected.

Doctor

DoctorID
Title
FirstName
Surname
Qualification

Patient (1)

PatientID
Title
FirstName
Surname
Address
Telephone
DoctorID *

Flat File

A flat file is a database that consists of a single file. The flat file will most likely be based around a single entity and its attributes.
Flat files are typically written out in the following way:

Entity1(Attribute1, Attribute2, Attribute3 …)

The example in the table shows the entity Car. Age and Price are attributes of each car. For this example, the description would be laid out as:

Car(CarID, Age, Price)

Car			(2)
CarID	**Age**	**Price**	
Car1	5 years	£1,500	
Car2	2 years	£2,400	

Primary Key

A primary key is a unique identifier for each record in the table. In example (2), the unique identifier is the CarID as this is always different for each row in the table. In example (1), the primary key for the Doctor table would be DoctorID, and the primary key for the Patient table would be PatientID. The primary key is shown by underlining it.

Foreign Key

A foreign key is the attribute which links two tables together. The foreign key will exist in one table as the primary key and act as the foreign key in another. In example (1), DoctorID is the foreign key in the Patient table. The foreign key is shown using an asterisk.

Secondary Key

A secondary key allows a database to be searched quickly. Looking at example (1), the patient is unlikely to remember their patientID but will know their surname. Therefore, a secondary index (secondary key) is set up on the surname attribute. This makes it possible to order and search by surname which makes it easier to find specific patients in the database.

Entity Relationship Modelling

When relating two databases together there are different degrees of relationships they can possess:

- One-to-one: Each entity can only be linked to one other entity, such as the relationship between a husband and wife. The husband entity can only be associated with one wife entity and vice versa.
- One-to-many: One table can be associated with many other tables, such as a mother having multiple children. Similarly, multiple child entities can be linked to the same mother entity.
- Many-to-many: One entity can be associated with many other entities and the same applies the other way round. An example is students and courses - each student can enrol in more than one course and each course can have more than one student.

One-to-one relationships are demonstrated using a single line used to connect two entities. A one-to-many relationship will have a branch on one side, while a many-to-many relationship has branches on both sides.

To practice entity relationship modelling, think about different real-world objects and how they are linked to each other. Consider, for example, the link between customers, orders and products.

Handling Data

Capturing Data

Data needs to be input into the database and there are multiple methods of doing this. The chosen method is always dependent on the context. For example, if pedestrians are participating in a survey, their responses will need to be manually entered into the database.

Data is also captured when people pay cheques. Banks scan cheques using Magnetic Ink Character Recognition (MICR). All of the details excluding the amount are printed in a

special magnetic ink which can be recognised by a computer but the amount must be entered manually. Optical Mark Recognition (OMR) is used for multiple choice questions on a test. Other forms use Optical Character Recognition (OCR).

Selecting and Managing Data
Selecting the correct data is an important part of data preprocessing. This could involve only selecting data that fits a certain criteria to reduce the volume of input. For example, a camera catching speeding cars will only select cars going above a certain speed. Then, background information will be removed so only the number plate is added to a record. Collected data can alternatively be managed using SQL to sort, restructure and select certain sections.

Exchanging Data
Exchanging data is the process of transferring the collected data. One common way of exchanging data is EDI (Electronic Data Interchange). This doesn't require human interaction and enables data transfer from one computer to another.