Oxford excellence for the Caribbean

Information Technology

THIRD EDITION

CSEC®

FOR THE NEW SYLLABUS

Sample pages

with online support

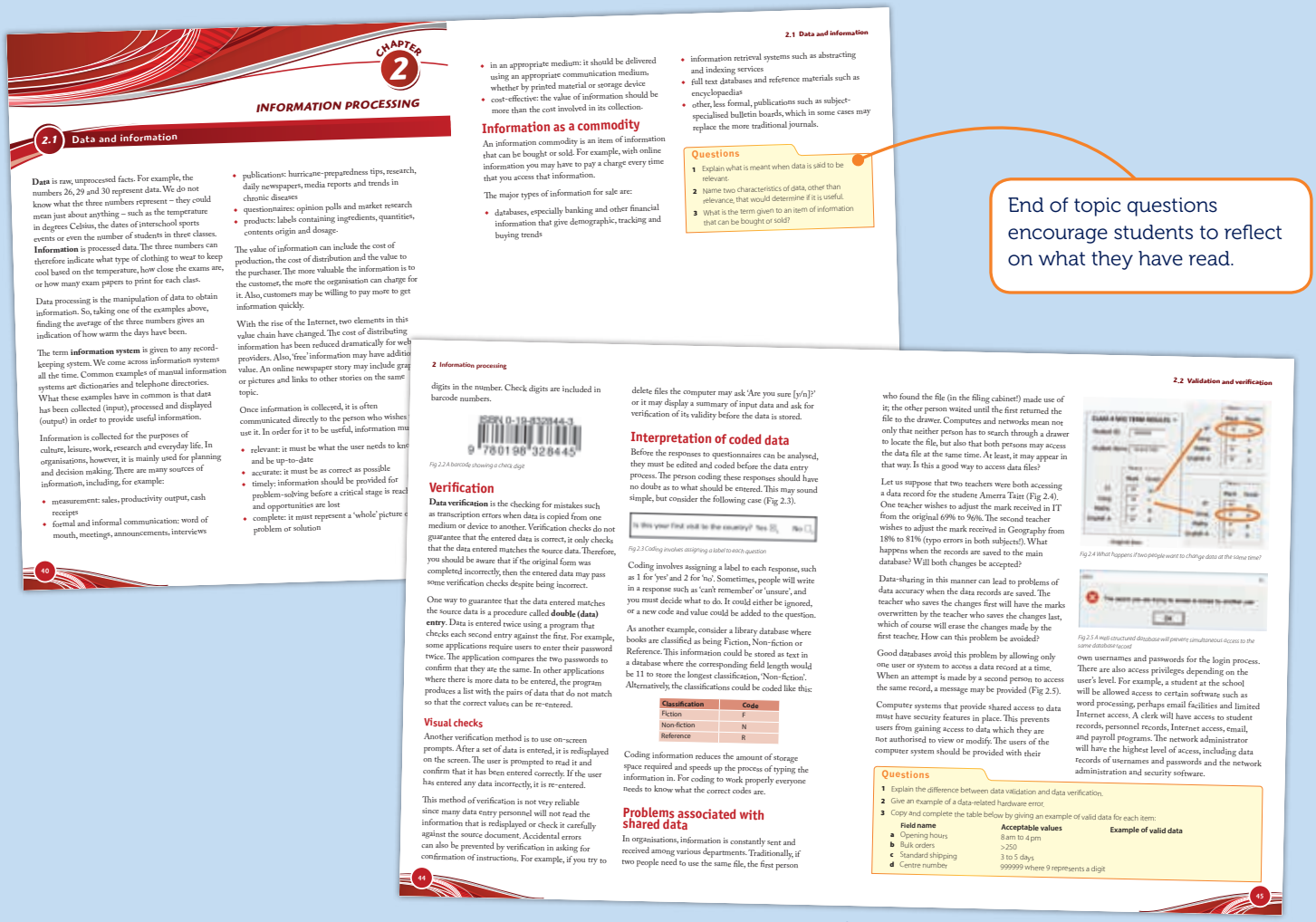
Glenda Gay
Ronald Blades

OXFORD

Instil confidence and support success in Information Technology

Oxford Information Technology for CSEC® is the market leading textbook for Information Technology providing comprehensive coverage of the updated CSEC® syllabus and essential support for the School-Based Assessment. The third edition focuses on developing students' problem solving skills through the use of ICT tools, helping them to embrace new technologies and fully preparing them with the knowledge and expertise to achieve in the exam.

- **Up-to-date content with Caribbean examples** makes learning more relatable.
- **Topic discussions, worked examples, and a wide range of questions** including exam-style questions provide extensive support and practice to prepare students for the exam.
- **Comprehensive support for the School-Based Assessment** with hints and guidance, and practice questions makes this practical aspect of the qualification more accessible to students, boosting their confidence.



Sample pages from IT for CSEC® Student Book, Unit 2

Contents

2.1 Data and information

Data is raw, unprocessed facts. For example, the numbers 26, 29 and 30 represent data. We do not know what the three numbers represent – they could mean just about anything – such as the temperature in degrees Celsius, the dates of interschool sports events or even the number of students in three classes. **Information** is processed data. The three numbers can therefore indicate what type of clothing to wear to keep cool based on the temperature, how close the exams are, or how many exam papers to print for each class.

Data processing is the manipulation of data to obtain information. So, taking one of the examples above, finding the average of the three numbers gives an indication of how warm the days have been.

The term **information system** is given to any record-keeping system. We come across information systems all the time. Common examples of manual information systems are dictionaries and telephone directories. What these examples have in common is that data has been collected (input), processed and displayed (output) in order to provide useful information.

Information is collected for the purposes of culture, leisure, work, research and everyday life. In organisations, however, it is mainly used for planning and decision making. There are many sources of information, including, for example:

- measurement: sales, productivity output, cash receipts
- formal and informal communication: word of mouth, meetings, announcements, interviews
- publications: hurricane-preparedness tips, research, daily newspapers, media reports and trends in chronic diseases
- questionnaires: opinion polls and market research
- products: labels containing ingredients, quantities, contents origin and dosage.

The value of information can include the cost of production, the cost of distribution and the value to the purchaser. The more valuable the information is to the customer, the more the organisation can charge for it. Also, customers may be willing to pay more to get information quickly.

With the rise of the Internet, two elements in this value chain have changed. The cost of distributing information has been reduced dramatically for web providers. Also, 'free' information may have additional value. An online newspaper story may include graphs or pictures and links to other stories on the same topic.

Once information is collected, it is often communicated directly to the person who wishes to use it. In order for it to be useful, information must be:

- relevant: it must be what the user needs to know, and be up-to-date
- accurate: it must be as correct as possible
- timely: information should be provided for problem-solving before a critical stage is reached and opportunities are lost
- complete: it must represent a 'whole' picture of a problem or solution

- in an appropriate medium: it should be delivered using an appropriate communication medium, whether by printed material or storage device
- cost-effective: the value of information should be more than the cost involved in its collection.

Information as a commodity

An information commodity is an item of information that can be bought or sold. For example, with online information you may have to pay a charge every time that you access that information.

The major types of information for sale are:

- databases, especially banking and other financial information that give demographic, tracking and buying trends
- information retrieval systems such as abstracting and indexing services
- full text databases and reference materials such as encyclopaedias
- other, less formal, publications such as subject-specialised bulletin boards, which in some cases may replace the more traditional journals.

Questions

1 Explain what is meant when data is said to be relevant.

2 Name two characteristics of data, other than relevance, that would determine if it is useful.

3 What is the term given to an item of information that can be bought or sold?

A common problem with manually entering data into a computer system is that it is very easy to input incorrect data. Examples include users mistyping a name or address in a database so that letters are sent to the wrong person. However, it is also possible for a device to misread a barcode and therefore not provide any information to the user.

Errors

Errors occur in any computer information system. There are several approaches to dealing with the problems that errors cause, specifically preventing errors (so that they do not occur) and detecting when errors do occur (so that they can be corrected). Below a few approaches are considered.

Data entry errors

Sometimes data cannot be automatically scanned into the system for storage, so a common way is to type the data directly into the computer system. Data entry errors, such as transposition errors, would then occur when an operator enters data using a keyboard and mixes up digits and/or letters. For example, the number 32 may be entered as 23. Other errors occur when data entry requirements are not clear. Do you interpret 07/01/2019 as July 1 2019 or 7 January 2019? These data entry errors can be either accidental or deliberate.

- Accidental errors occur unintentionally: data is entered or a command is issued by mistake, in good faith, but in error.
- Deliberate: if errors are made deliberately, then the user probably knows enough to get around any validation checks. Examples include a disgruntled employee entering fictitious data into a company's database. Possible solutions to this problem include file access permissions imposed by the operating system, better vigilance of the administrator and the cross-referencing of data with other supposedly reliable sources.

Software and hardware errors

There are two types of error here. Software might malfunction, erasing or corrupting previously entered data. Hardware may develop a fault (often intermittent) that corrupts data. Examples include bad sectors on a hard disk, bad memory or a power surge. The application may appear to accept data correctly but when that data is retrieved it is corrupt.

Transmission errors

Transmission errors occur when data received is not the same as that transmitted by the sender (Fig 2.1).

Fig 2.1 Illustration of transmission errors

Validation

Data validation is the computerised checking of input data for errors (data that may be unreasonable or incomplete) before it is processed. It does not confirm the accuracy of data. There are different methods of validation. The most appropriate method(s) to use will depend upon what data is being entered.

Range check

Range check ensures that the data entered is within a certain range. For example, when you enter a number corresponding to a month of the year, the range of acceptable numbers is 1 to 12.

Reasonableness check

This check tests whether the data obeys specified criteria. For example, the age of a child at preschool could be about 4 years old, but not 14 years. Reasonableness checks are therefore used to detect doubtful data.

Data type check

This is also known as a character or alphanumeric check. When a database is created, each field will

accept a specific type of data. Whenever data is entered into a field the database will check that it is of the correct type, for example alphabetic or numeric. If it is not, then an error message will be displayed, and the data will have to be re-entered.

Table 2.1 *Examples of valid and invalid data types*

Field name	Type	Valid data	Invalid data
Date of birth	Date	19/10/1994	19/19/94, or 23
Percentage mark	Numeric	56, 99	A+, Pass, 125, –15

Notice that a type check is not a very good validation check. There are many entries you could put in the 'Valid data' column in Table 2.1 that would pass the type check but are clearly incorrect.

The data type check is particularly important if a fixed length field is defined to store the data. Any extra characters that exceed the maximum length would be lost. Length checks are usually only performed on alphabetic or alphanumeric data.

Table 2.2 *Examples of valid field lengths*

Field name	Maximum length	Valid data	Invalid data
Student ID	6	826025	82–60–45
Grade	2	B+, C	A++, Fail

Consistency check

A consistency check compares the contents of two or more fields to make sure that they make sense. It is also called an inconsistency check, since it mostly identifies errors and discrepancies in the data. This check compares new data with previously entered data. For example, checking that the age entered corresponds to the age calculation from the date of birth. Consider the following employee record:

Employment Status Form	
Are you unemployed?	No
Unemployment benefits claimed?	Yes
Number of years working	12
Current occupation	Foreman

It would be useful to have a consistency check to cross-check the information in the 'Are you unemployed?' and 'Unemployment benefits claimed?' fields, since you should not claim for unemployment benefits if you are still working. In this example, therefore, either an error has been made on input or this is a deliberate attempt to claim benefits while still working. Other examples of consistency checks are:

- Single mothers with children can claim for childcare allowance. Check that the number of children is not zero.
- Only full-time employees are paid overtime. Check that these fields correspond.

Presence check

This type of check ensures that the data is actually entered. It can be used on any field in a database. For example, every person in a particular database must be assigned to a department. Therefore, a presence check on each employee's record could ensure that a department is entered in the form.

Format check

This check verifies that the data has been entered in the correct format. For example, a national ID number may have the format 999-9999-X999, where 9 represents a number and X represents an alphabetic character.

Length check

This check verifies that the data entered is the correct length. For example, a password for a credit card may be four digits long, therefore entering three or five digits may result in an error.

Check digit

A check digit is an extra digit added to the end of a code. It is used to detect errors arising from transcription and also to ensure that codes originally produced by a computer are re-entered into another computer correctly. It is calculated from the other

digits in the number. Check digits are included in barcode numbers.

Fig 2.2 A barcode showing a check digit

Verification

Data verification is the checking for mistakes such as transcription errors when data is copied from one medium or device to another. Verification checks do not guarantee that the entered data is correct, it only checks that the data entered matches the source data. Therefore, you should be aware that if the original form was completed incorrectly, then the entered data may pass some verification checks despite being incorrect.

One way to guarantee that the data entered matches the source data is a procedure called **double (data) entry**. Data is entered twice using a program that checks each second entry against the first. For example, some applications require users to enter their password twice. The application compares the two passwords to confirm that they are the same. In other applications where there is more data to be entered, the program produces a list with the pairs of data that do not match so that the correct values can be re-entered.

Visual checks

Another verification method is to use on-screen prompts. After a set of data is entered, it is redisplayed on the screen. The user is prompted to read it and confirm that it has been entered correctly. If the user has entered any data incorrectly, it is re-entered.

This method of verification is not very reliable since many data entry personnel will not read the information that is redisplayed or check it carefully against the source document. Accidental errors can also be prevented by verification in asking for confirmation of instructions. For example, if you try to delete files the computer may ask 'Are you sure [y/n]?' or it may display a summary of input data and ask for verification of its validity before the data is stored.

Interpretation of coded data

Before the responses to questionnaires can be analysed, they must be edited and coded before the data entry process. The person coding these responses should have no doubt as to what should be entered. This may sound simple, but consider the following case (Fig 2.3).

Is this your first visit to the country? Yes ☒$_1$ No ☐$_2$

Fig 2.3 Coding involves assigning a label to each question

Coding involves assigning a label to each response, such as 1 for 'yes' and 2 for 'no'. Sometimes, people will write in a response such as 'can't remember' or 'unsure', and you must decide what to do. It could either be ignored, or a new code and value could be added to the question.

As another example, consider a library database where books are classified as being Fiction, Non-fiction or Reference. This information could be stored as text in a database where the corresponding field length would be 11 to store the longest classification, 'Non-fiction'. Alternatively, the classifications could be coded like this:

Classification	Code
Fiction	F
Non-fiction	N
Reference	R

Coding information reduces the amount of storage space required and speeds up the process of typing the information in. For coding to work properly everyone needs to know what the correct codes are.

Problems associated with shared data

In organisations, information is constantly sent and received among various departments. Traditionally, if two people need to use the same file, the first person

who found the file (in the filing cabinet!) made use of it; the other person waited until the first returned the file to the drawer. Computers and networks mean not only that neither person has to search through a drawer to locate the file, but also that both persons may access the data file at the same time. At least, it may appear in that way. Is this a good way to access data files?

Let us suppose that two teachers were both accessing a data record for the student Amerra Taitt (Fig 2.4). One teacher wishes to adjust the mark received in IT from the original 69% to 96%. The second teacher wishes to adjust the mark received in Geography from 18% to 81% (typo errors in both subjects!). What happens when the records are saved to the main database? Will both changes be accepted?

Data-sharing in this manner can lead to problems of data accuracy when the data records are saved. The teacher who saves the changes first will have the marks overwritten by the teacher who saves the changes last, which of course will erase the changes made by the first teacher. How can this problem be avoided?

Good databases avoid this problem by allowing only one user or system to access a data record at a time. When an attempt is made by a second person to access the same record, a message may be provided (Fig 2.5).

Computer systems that provide shared access to data must have security features in place. This prevents users from gaining access to data which they are not authorised to view or modify. The users of the computer system should be provided with their own usernames and passwords for the login process. There are also access privileges depending on the user's level. For example, a student at the school will be allowed access to certain software such as word processing, perhaps email facilities and limited Internet access. A clerk will have access to student records, personnel records, Internet access, email, and payroll programs. The network administrator will have the highest level of access, including data records of usernames and passwords and the network administration and security software.

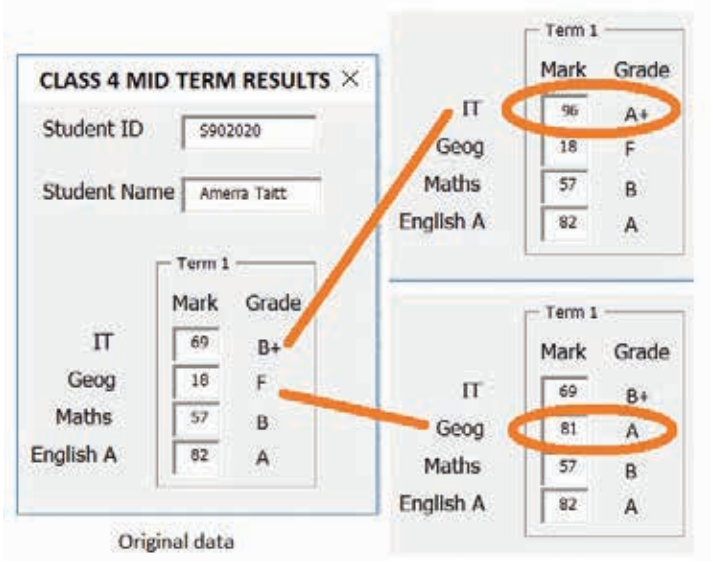


Fig 2.4 What happens if two people want to change data at the same time?

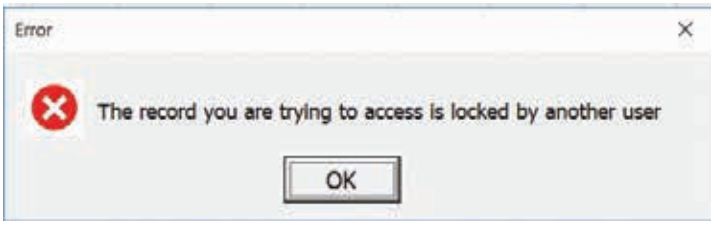


Fig 2.5 A well-structured database will prevent simultaneous access to the same database record

Questions

1 Explain the difference between data validation and data verification.

2 Give an example of a data-related hardware error.

3 Copy and complete the table below by giving an example of valid data for each item:

	Field name	Acceptable values	Example of valid data
a	Opening hours	8 am to 4 pm	
b	Bulk orders	>250	
c	Standard shipping	3 to 5 days	
d	Centre number	999999 where 9 represents a digit	

Many of the output devices discussed in Chapter 1 are human-readable, meaning that a hard copy of the output is printed as reports, graphs, charts and so on. Other devices are machine-readable, meaning that the output is in a form that only a computer can process. An example is output to a monitor. This is called soft-copy output since it is not printed on paper or other physical material.

Data-capture forms

These forms must be designed so that their instructions are clear and concise, leaving no doubt as to how to enter the data onto the form. For example, the format for dates could be specified as 'dd/mm/yyyy' so that the format 'mm/dd/yy' is not used in error.

The responses from the forms are called human-readable since data entry personnel manually enter the responses written on the forms. This method of data capture is liable to transcription and other errors when the data is entered into the computer. One method of avoiding transcription errors is double (data) entry, where the data is entered twice by two different people and the computer will only accept the data if the two versions are identical.

Turnaround document

A turnaround document is a machine-readable document that has some information printed on it by a computer but has more information added to it by a human. It is then fed back into a computer to transfer this newly added information. These documents serve two purposes. They are used to:

- verify the accuracy and completeness of information that has already been entered
- update information already entered with additional data.

Optical mark recognition and optical character recognition are often used together in a turnaround document. Figure 2.6 shows an example of a turnaround document used to record the reading on an electricity meter.

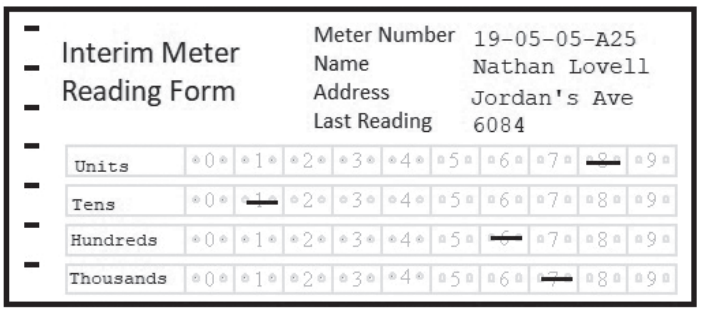


Fig 2.6 The meter reader has marked the reading as 7618 on this turnaround document

Another example of a turnaround document is the multiple-choice sheets used in examinations. Information such as the candidate number, subject and school code is printed on the multiple-choice answer sheet by the computer. The student takes the test and fills in the answer grid by making marks in the appropriate boxes using a pencil or ballpoint pen. The form is then returned to the examination council to be fed into a special reader.

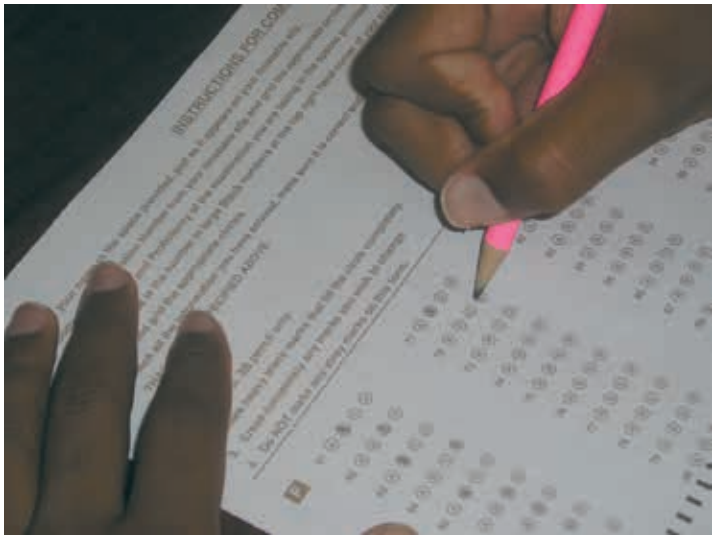


Fig 2.7 Multiple-choice exam sheet

Optical character recognition (OCR) is used to read the student's pre-printed information from the form, whereas optical mark recognition (OMR) is used to read the answers that have been added. All the information can be read into the computer automatically. The sheet is marked, and the total mark is printed without any need for human intervention.

Turnaround documents allow cheap, fast input of information into a computer system. Data can also be printed on turnaround documents in barcode format.

Alternatives to turnaround documents

In some industries, turnaround documents are being replaced by small handheld computers, including mobile devices. These pocket-sized devices have a display screen and touchscreen keyboard. For example, many local utility companies now issue their meter readers with handheld computers. At the start of the day the names and addresses of the customers to be visited are downloaded into the computer. As the meter reader visits customers' homes, he or she types the meter readings into the computer. At the end of the day the readings can be transferred automatically onto the main computer to issue bills. Another example involves courier services that equip drivers with these devices. As the courier delivers a package, the recipient uses a stylus to sign the delivery document on the device. This signature is captured digitally and uploaded to the database as proof of delivery.

Fig 2.8 Turnaround documents are being replaced by small handheld computers

Data logging

Computers are often used by companies and scientists to automatically measure and record changes in conditions such as the temperature, the speed at which a ball is travelling in sports such as cricket and baseball, the amount of light or oxygen in a room or even the level of noise being made by vehicular traffic near a hospital.

Data logging is a method of automatic data capture where a reading from a sensor is input at regular intervals. This data can then be processed to provide analysis of the environment (Fig 2.9).

Fig 2.9 This student is checking water quality using a sensor connected to a data logger. This data can then be processed to provide analysis of chemicals in the water

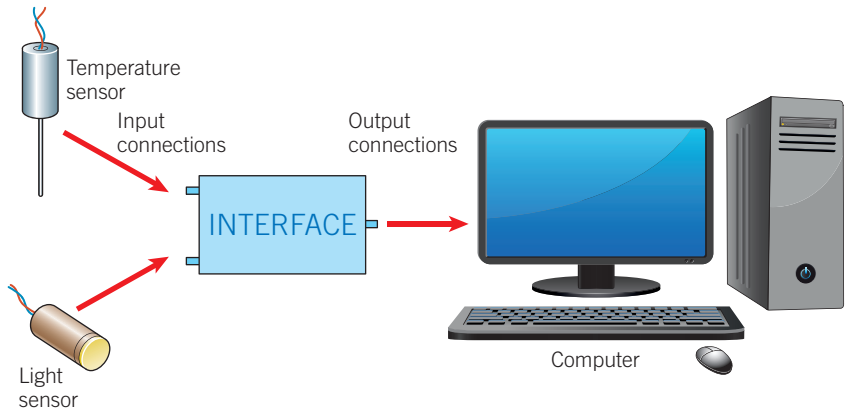


Fig 2.10 Components of a data logging system

A data logging system consists of sensors connected via an interface to a computer and some data logging software (Fig 2.10). The sensors will take measurements and at required intervals the software will record the data. The results can then be displayed as a graph or a table.

Microfilm

Microfilm is often used as an alternative to the printer. The output is 'printed' on a roll of film (computer output microfilm) or sheet of film (computer output microfiche) rather than paper. In addition to being faster, this method also condenses large stacks of paper into small amounts of microfilm or microfiche with no special programming. The drawback of computer output microfilm or microfiche (COM) is that it takes a special device to print the microfilm and a special viewer to read it.

Questions

1 State which of the following output is hard-copy output or soft-copy output:
   a human-readable
   b machine-readable.

2 Give an example of a document that is machine-readable.

File organisation and access relates to the use of records, fields and files.

- A **field** contains a single data item, and many fields comprise a record. Each field has a name and one is the key field used to identify the record.
- A **record** is a collection of related data fields (possibly of different data types) and treated as a single item for processing.
- A data file is a collection of records holding the same type of information but about different objects or individuals.

A file has three important characteristics:

1 It can be permanent or temporary.
2 The records of the file are specially organised on the secondary storage device. This is called file organisation.
3 Records are accessed (or located) using different methods.

Master and transaction files

Many businesses and organisations regularly access, modify and store large amounts of files. These files are given special names to identify their purpose. A master file is a permanent file which is kept up-to-date. It stores the main information, summary data and key fields in the data.

The master file contains two types of data:

- permanent data, such as employee personal data, payroll data employee status and job title
- temporary data, which is updated on a regular basis, such as hours worked and taxes deducted.

A **transaction file** is a temporary file which is used to update the master file after a certain time (at the end of each day or week, for example). The transaction file updates the master file. The records in the transaction file are used to perform three important operations:

- add: put a new record into the master file
- update: change the contents of a record or a field that already exists
- delete: remove a record from the master file.

Record matching

A **primary key** is normally used to identify the record you want to update or delete. It is a field in the record whose value is unique to that record. For instance, in a student record, the Student ID is normally used as the key field. Without a key field to identify the record you want you cannot delete or update records.

To delete or update records in a master file, compare the primary key in the transaction record with that in the master file record. If there is a match, you can update or delete the master file record. If both files are ordered (sorted) on the key field, then this record matching operation functions correctly, but if either the transaction or the master file is unordered, record matching cannot work.

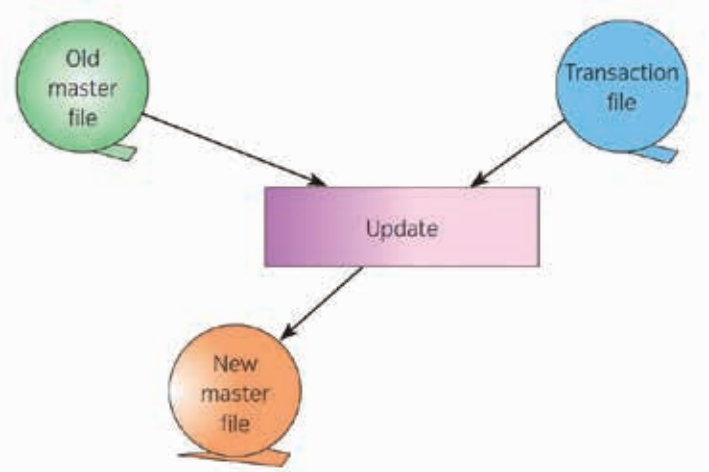


Fig 2.11 Updating a master file with a transaction file

There are three possibilities when updating a master file:

1 Transaction record key is *less than* the master file record key.
The transaction record is added to the master file.
Next transaction record is read.

2 Transaction record key is *equal to* the master file record key.
The master record is deleted or updated.
Next transaction record is read.
3 Transaction record key is *greater than* the master record key.
No more transactions for this master record.
Write master record to new master file.
Read next master file record.

Processing errors

Table 2.3 lists some errors or situations which can occur during processing.

Table 2.3 *Error messages during file processing*

Situation	Error
If the master file encounters an 'end of file' marker in the transaction file, then there is no data on which to update the master file	No data found
If you are searching for a record to update or delete in the master file, and it is not found	Record does not exist or invalid primary key in the transaction record. If you reach the end of the master file before the end of the transaction file, then either you are adding new records in the transaction file to the master file, or you are not updating the most current master file, or your master file may be corrupt
If you are searching for a location to add a record in the master file, and you find the record	Trying to add a record that already exists

Serial and sequential file organisation

Serial file organisation is the simplest type of file organisation. The records are arranged one after another, in the order in which they were added. That is, they have not been sorted into any particular order. Examples of serial files (also called unordered files) include unsorted invoices for customers for the day, and a typical shopping list. Serial files can be stored on tape, disk or in memory. A serial file is used mainly for backup purposes. Recording data in the order in which transactions are made is also useful if you need to restore transactions that have been lost from the master file through hardware or other problems.

Unordered file	Ordered file
Record M	Record A
Record H	Record B
Record B	Record G
Record N	Record H

A sequential file is one in which the records are stored in sorted order on one or more key fields. Examples of sequential files include invoices for customers sorted on customer number, and class registers sorted on last name. Magnetic tape is sequential by its very nature, just like listening to a cassette tape or watching a movie. To access a particular section or continue from where you left off, you must start at the beginning and scan forward until you reach the specific one.

Searching for a record

To search for a particular record, all the preceding records must be read. The main drawback to inserting, deleting or amending records in both serial and sequential files is that the entire file must be read and then the records written to a new file. Since disk access is slow, computer time is wasted even if only a few records are involved. For instance, if 10 records are to be added into a 1000-record file, then 1000 records will have to be read from the old master file and after the 10 insertions from the transaction file, 1010 records are written to the new master file. It therefore takes a long time to insert a new record with serial organisation. To maximise efficiency of processing, use sequential organisation, where the records are arranged in order by the value of the key field common to all records.

Summary of reading, writing and sorting for sequential files

Read access

- Records are read from the beginning of the file until the desired item is found
- If accessing a record, access time is increased to read the entire file or many records but decreased if few records are to be accessed

Write access

- Adding records to the end of the file is easy (may require sorting). However, it is difficult to add or delete data in the middle of the file.

Sorting

- Sequential files are often sorted on the record key to make processing simpler. However, sorting can be time consuming for large files.

Adding a record

For serial files, you simply add the new record to the end of the file. However, the major purpose of sequential files is to preserve the ordering of the file. This means that the record must be inserted into the file in the correct position and not at the end of the file as with serial files. Also, you cannot just insert all changes to records in sequential files into the existing file – you must create a new file that contains the inserted records.

The algorithm for adding a record is:

1 All the records with a key value less than the record to be inserted are read and then written to the new file.
2 The record to be inserted is written to the new file.
3 Finally, the remaining records are written to the new file.

Updating a record

Updates are normally done using magnetic tape in batch mode. All the updates are gathered together into a transaction file, and then applied to the master file together. Updating, therefore, is again accomplished by creating a new file which contains the updated records. Sorting the master file and transaction file records in the same order improves the efficiency of the updating process.

To amend a record in a sequential file:

1 All the records with a key value less than the record to be amended are read and then written to the new file.
2 The record to be updated is read, any changes are applied to it and the updated record is written to the new file.
3 Finally, all the remaining records are written to the new file.

Deleting a record

With both types of files, the only way to delete records is to create a new file which omits the records marked for deletion.

To delete a record in a serial file:

1 Compare each record with a key value of the record to be deleted (since the transaction file is not sorted).
2 If it is not the record to be deleted, then write that record to the new file.
3 When the record to be deleted is encountered it is not written to the new file.
4 Finally, all the remaining records are written to the new file.

To delete a record in a sequential file:

1 All the records with a key value less than the record to be deleted are written to the new file.
2 When the record to be deleted is encountered it is not written to the new file.
3 Finally, all the remaining records are written to the new file.

Direct access file organisation

A direct access file, also called a random access file, allows access to a particular record in the file using a key. This makes it much easier to find, read, delete, update and insert records. The file is organised like a one-dimensional table on disk where each record is a

part of the table. The record number acts like a table index to allow you to find the records.

You can access a record directly or randomly by calculating its location using a mathematical formula and going directly to the record. For example, when you input an ID number, the mathematical formula uses it to produce a value that points to the storage location on disk where the record can be found.

Direct access files also support sequential access by allowing the records to be read one after another. The records in a direct access file are not sorted in ascending or descending order, but are stored in random order.

With hard disks, direct access is possible. With an audio compact disk, for example, you can play the songs in random order or go directly to the track you want to hear. However, not only must the medium allow for random access to records, but the file itself must be organised so that you can go directly to the record you want to retrieve. This can be compared to sequential organisation, as on a magnetic tape. You have to start at the beginning and run the tape forward until you get to the song you want to hear.

Summary of direct access file organisation

This organisation is best for:

- files which seldom change in size
- files which require frequent updates
- single record enquiries and updates
- processes which require fast access to records
- storage of master file records on direct access media only (such as hard disks)
- accessing disk file records sequentially or directly.

If processed directly, they need not be processed in order.

Index sequential file organisation

An **indexed file** is used to speed up the key search in a file. You can think of it as a one-column table organised in ascending order and stored on disk. The primary key in the table is used as an index to the record. It is just like the index of a book where the key value (topic) has a pointer to the storage location (page number) where the information is stored.

Many applications require a mix of sequential and direct processing of records. Consider a file containing customer accounts with three fields: 'Account Number', 'Credit limit' and 'Balance'. Every time a customer wants to make a purchase, his or her credit limit and balance must be checked; this requires individual access to his or her record. Every month, a statement must be produced for each customer; this requires sequential access to the whole file.

An index is a special file of records with two attributes: record key and the storage address of the corresponding record in the indexed file. A partial index containing the highest or the lowest key value in each block of records is useful when the index itself is organised sequentially. Ideally, you want to have the index in main memory. Then you can search the index quickly to obtain the storage address, and then retrieve the required record in a single disk access.

Searching for a record

The search key is compared with the index keys to find the highest index key that comes before the one you are searching for. Then a linear search is performed from there onward, until the search key is matched or until the record pointed to by the next index entry is reached.

For example, suppose you are in a supermarket where the items are stored on the shelves in alphabetical order (Fig 2.12). You wish to find forks. Instead of walking up and down each aisle, you look up at the listing (index) of what is in each aisle. This tells you which aisle to walk down to find the forks. You know to stop looking once you have found forks on the listing (Fig 2.13).

In spite of the double file access (index + data) needed by this kind of search, the decrease in access time with respect to a sequential file is significant.

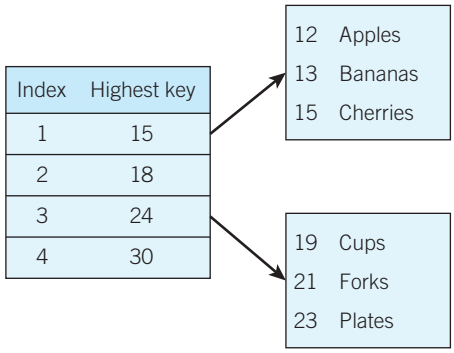


Fig 2.12 Index sequential file access

Fig 2.13 Supermarket aisle signs are similar to index sequential searches

Summary of index sequential access file organisation

- Instead of having an index entry for each record, have an entry for each block of records starting from the lowest or highest record.
- Leave spaces in each block to allow for easy insertions.

Table 2.4 summarises methods of access to a record for a variety of file structures.

Table 2.4 *Searching for a record can be achieved through various file organisations*

File structure name	Structure details	Access method
Serial file	Unordered records	Sequential access
Sequential file	Ordered records	Sequential access
Partially indexed file	Ordered records	Sequential access to index, followed by direct access to first record in the group, then sequential access to find the desired record
Fully indexed file	Unordered records	Sequential access to the index, followed by direct access to the data file
Direct access file	Unordered or ordered records	A calculation provides the address (location) of a record, followed by direct access to the record

Questions

1 Write the connection between a field, record and file.

2 Why are master and transaction files needed in most businesses?

3 What is the special name given to the field that normally identifies a record?

4 State the type of file organisation for each of the following descriptions:

- **a** records arranged one after another, in the order in which they were added
- **b** records stored in sorted order
- **c** records stored in random order
- **d** records stored with directory-type listings to denote location.

The processing of data into information occurs when a machine or processor acts on the input it receives. An example of this is an electric kettle which senses when the water temperature reaches boiling point and switches off. A car production line can sense when a car body is in a certain position and then act to weld together the relevant parts of a car. **Information processing** is only valuable if the information can be stored and retrieved quickly, accurately and efficiently and cannot be changed accidentally.

Setting up an information processing system

Before setting up an information processing system, a business should consider the following questions:

- Will computerisation really solve the particular problem?
- Is it cost-effective in the long term?
- How large is the amount of data to be handled?
- Is high processing speed really important?
- Can the present staff manage the system?
- Will the changes caused by computerisation lower the morale of staff?
- How can the loss of jobs be handled properly?
- What can be done to help staff adapt to computerisation?

Examples of information processing

Information processing can be done in almost all sectors of business. It is also becoming increasingly popular at home where bills can be paid either by phone or on the Internet. Other people use information processing for research and education through online classes where assignments and reports can be submitted electronically. In fact, anyone can use the Internet to find a wealth of information such as current affairs news, stock prices, educational materials, online banking and investments, shopping for goods and services, communication, and the exchange of information with other people around the world.

Table 2.5 *Advantages and disadvantages of information processing*

Advantages	Disadvantages
◆ Tasks can be completed faster because data and information can be processed at amazing speeds. ◆ Computer storage devices can store enormous amounts of data and information for future use. ◆ Automation can be introduced. That is, tasks can be completed with little human intervention. ◆ Management can analyse new information and trends more quickly. ◆ Data and information can be shared with other computers.	◆ It may need a high initial investment in equipment and training. ◆ More money may be needed to employ specialised staff to operate and design the information processing system. ◆ Some jobs may be lost as a result of computerisation, which may lower the morale of staff members. ◆ Some staff must be trained or retrained. ◆ Face-to-face interaction between staff may be reduced.

Health care

Information processing in health care may be used to:

- maintain patient records in hospitals and clinics
- monitor patients' vital signs in hospital, and at home
- perform computer-assisted medical tests
- research and diagnose medical conditions
- operate implanted devices such as pacemakers which allow patients to live longer
- control surgical instruments during operations that require great precision, for example laser eye surgery and heart surgery
- enable 'telemedicine' through computers with video conferencing capabilities
- train surgeons before they perform surgery.

Banking

Computers are used to keep track of all bank transactions. Customer accounts need to be updated every time a payment transaction is made whether by cheque, card or **EFT (electronic funds transfer)** at the point of sale (EFTPOS) is useful for customers in a shop. The bank card is inserted into a reader attached to the point of sale (POS) terminal. The payment is then made directly from the customer's bank account to that of the shop. The procedure is as follows:

1. The cost of all of the items to be purchased is added up, usually on a computerised cash register.
2. The customer presents his or her debit or credit card to the shop assistant.
3. The card is inserted so the chip can be read or swiped through a magnetic strip reader to input the card number and expiry date to a computer.
4. The card number, payment amount and identity of the company that has sold the goods are sent to the bank's computer using a modem and telephone line.
5. The customer types a four-digit **personal identification number (PIN)** on the keypad and presses the enter key to continue the process.
6. The bank's computer looks up the customer's account details in an accounts database.
7. If the card is valid and the customer has enough money in his or her account then the payment is approved.
8. The money is transferred electronically from the customer's bank account to the company's bank account.

Sometimes the magnetic strip reader cannot automatically read the card number from the card. If the magnetic strip has been damaged, then the sales assistant can enter the card number using a small keypad.

Payroll

A payroll system uses an information processing system to calculate the wages of each employee, print out pay-slips and record the information for accounting purposes.

Fig 2.14 Using a bank card in a reader attached to a point of sale terminal

Payroll example

Input: This may come from a database of employees' details, such as salaries, pay rates, bonus rates if employees are paid by the hour, then timesheets would be used to input and validate the number of hours worked and number of hours overtime, possibly using OMR or OCR techniques.

Processing: Using a software application such as a spreadsheet or more complex accounting software the computer then needs to calculate the gross amount earned by each employee, any bonuses, any deductions such as tax, national insurance, etc. and hence the net amount earned by each employee.

Output: The computer would need to print pay-slips. Use an impact printer if the pay-slip is required in duplicate for the employee's signature, or a non-impact printer otherwise. Update the employee database using a database integrated with the accounting software. Output details of payments to a banker's automated clearing service to pay money directly into employees' bank accounts using electronic commerce. Print summary reports.

Library

While travelling to a library to borrow books may be a regular activity for many students, libraries have been adjusting their collections, services and environments for the digital world. Most university libraries provide online access for their staff and students via an

e-information portal. This allows users to search online databases for e-journals, e-books, and articles in the online, digital or e-library. In some cases, users can also search for and request a paper-based version of the article and then travel to collect it if it is available.

Library example

Input: After logging in to the online library, the user can search for an article in an online database by entering information such as a subject, the name of an author or the title of a journal or article.

Processing: The library's online databases would have files (tables) containing details about the books, magazines and journals in the library. The system uses the key terms entered to conduct the search.

Output: The e-library may show different options to view the texts. For example, if a book is available for online reading, then you may be able to download it as a PDF document or read it in your browser, with special e-reader software or a Kindle device. Some online content offers audio features for the visually impaired or is formatted to deliver book page images. Some online libraries also provide links to other e-libraries and may indicate if a hard copy is available.

Control systems

You are surrounded by computer control systems but probably do not know it. Here are some examples of control systems.

- Traffic lights are triggered by movement sensors or the vehicle interacting with the sensor embedded in the road (Fig 2.15). A traffic light control system would not be very useful or safe if it did not respond adequately to the oncoming vehicles and stop the traffic! To do this, there has to be a computer program which is constantly looking at the data from the sensors and making decisions about what the output device (the traffic lights) should do.
- In the kitchen, microwave ovens, washing machines and tumble dryers all have control systems inside them to make them do their job at the press of a button. In the sitting room, remote-control televisions, video recorders and audio systems have built-in control systems.
- All modern cars have a management system which tells the engine what to do. This can control the flow of fuel to the engine and stops the engine from going too fast. Remote-control locks respond to a signal from the key to operate the locks on the doors.
- Buildings with air conditioning have sensors which detect the temperature and humidity inside the buildings and turn the heating or air conditioning on or off when needed.

Fig 2.15 Control systems are around us all the time

In these examples many of the sensors have digital outputs. In the traffic lights example, the embedded sensor that detects an approaching vehicle needs only to know whether a vehicle is there or not. So an on/off digital sensor is adequate. A digital temperature sensor would only be able to tell the system that the environment is hot or cold. It would not be able to tell the system how hot or how cold it is.

The processing of information is usually integrated into a control system which has sensors to input information, a processing unit (computer) which decides how to respond to the inputs, and output devices which do what is required. A control system also needs an interface unit between the computer and the input sensors and the output devices. This unit turns the signals from the sensors into something that the computer can understand, and the signals from the computer into something that will

work the output devices. The purpose of the interface is to make all parts of the system work with each other.

The processing unit in a control system may be a computer which has a program built into it. It will usually not be like the computers that you use, where you change the program by loading a new program from disk. The processing unit has a resident program in its electronic circuits. Such a system is called an **embedded controller**. Embedded controllers have only one program in their electronics to do the job they were designed to do. This makes them much cheaper to make because they do not need disk storage devices, a keyboard, a mouse or a screen. Once the system has been tested to make sure that it behaves as it should, these elements are not needed. The outputs will be the things that are being controlled, not a screen.

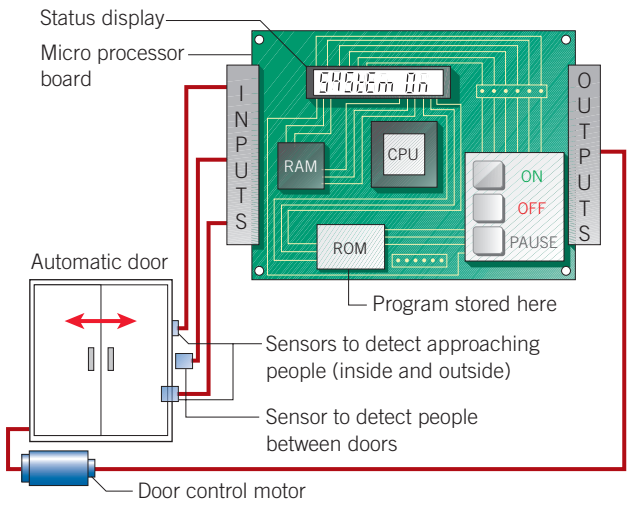


Fig 2.16 A typical control system looks like this. Note that it does not have a screen, disk or keyboard. The program is stored in ROM

Industry

Sometimes, in industry, a whole manufacturing process is controlled automatically by a computer system. Computer manufacturing systems do many tedious and repetitive tasks. In electronic circuit production, the components are automatically put in the right place on the circuit boards and then automatically soldered into position. The speed and accuracy at which these machines work are greater than those of skilled workers. The danger to the workers from the heat generated by soldering is removed by having automatic soldering systems.

The number of people needed in such areas of work has changed and usually been reduced. The skills the workers need have also changed. Rather than doing the repetitive tasks themselves, they now keep the machinery running and monitor activities.

Industry example

Input: Sensors take readings at regular time intervals and send the readings to the computer. Sensors may be measuring temperature, pressure, liquid flow rate and so on.

Processing: The computer analyses the readings and decides whether action needs to be taken.

Output: The computer sends output signals to devices which manage the process to increase pressure or temperature, for example. Some systems use an actuator, which is a device that reacts to a computer signal and operates a simple device such as a tap, motor or switch to regulate liquid flow.

Most of these systems use feedback, where the output affects the input.

Weather forecasting

Some of the world's most powerful computers are used to forecast the weather, which improves the accuracy of forecasts. People who rely on these forecasts include television companies, shipping companies, farmers, the military and outdoor sports organisations. Computer systems are also used to track hurricanes and tornados, monitor global warming, and monitor the ocean for systems of currents such as El Niño. Automatic data recording for weather forecasting has several advantages:

- It is more accurate than manual data collection.
- Computer data can be collected continuously whereas humans may get tired, and it can also be collected in situations not safe for humans.
- It is extremely fast. Computers can easily take thousands of measurements in a second. This means that events which could not be measured by a person can now be recorded for analysis later.

Table 2.6 *Sensors and their applications*

Sensor	Quantity measured	Application
Temperature sensor	How hot/cold it is	Monitoring the temperature in an oven
Light sensor	How light/dark it is	Turning street lights on when it is dark
pH sensor	The acidity of a liquid	Monitoring water pollution
Proximity sensor	Detects how close an object is to another object	Detecting how close a vehicle is when its driver is reversing near to a wall
Pressure pad	If a pad is being pressed	Detecting cars arriving at traffic lights
Button	If the button is being pressed	Obtaining a ticket to a paid car park
Light gate	Detects an object passing through the gate	Measuring the speed or acceleration of objects
Passive infrared (PIR)	Detects when a warm object moves into an area	Activating a burglar alarm if someone enters a room

Weather forecasting example

Input: Millions of pieces of data (observations such as temperature, pressure, humidity, infrared radiation) are collected from satellites, weather stations, weather balloons, aircraft, radar, weather ships and automatic weather buoys. All these readings are sent to the respective meteorological office's computer systems.

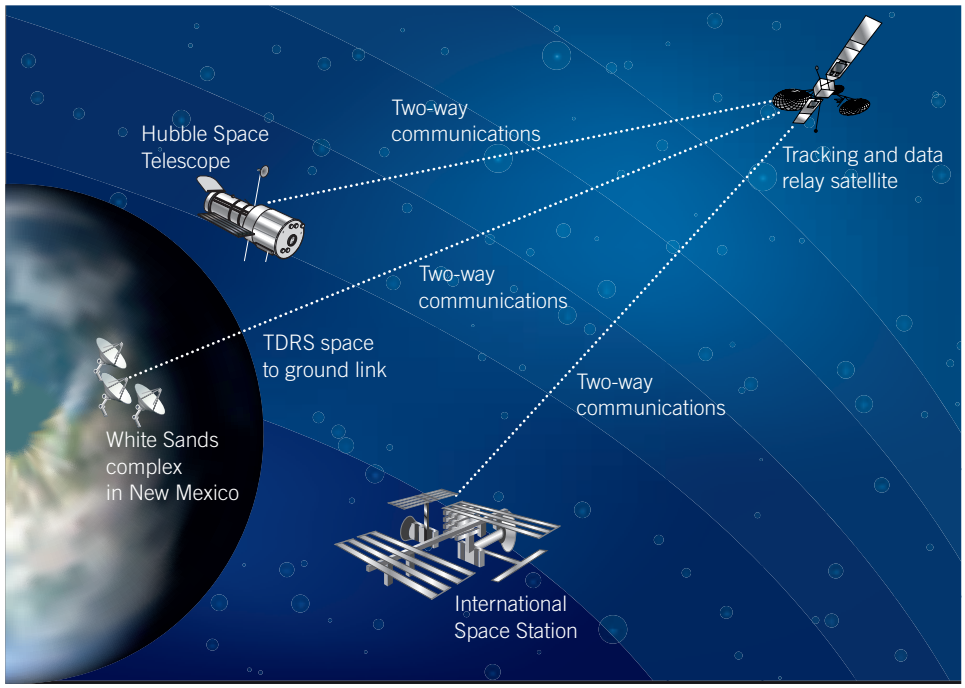


Fig 2.17 Data is collected from satellites orbiting the earth

Processing: The data is stored in a large database. The first task is to perform a quality control check on the data (validation) and to reject all invalid readings.

The data is formatted to fit in with a numerical model of readings. From this computer model, forecasts can be made. The bulk of processing is 'number-crunching' and solving thousands of inter-related equations.

Output: The forecasts are normally produced as global and local charts of weather information.

Supermarket stock control

Most businesses need to hold stocks of goods. Shops need to hold stocks of goods they sell, and manufacturers need to hold stocks of raw materials and

finished goods they make. The task of recording and maintaining stock levels is called stock control.

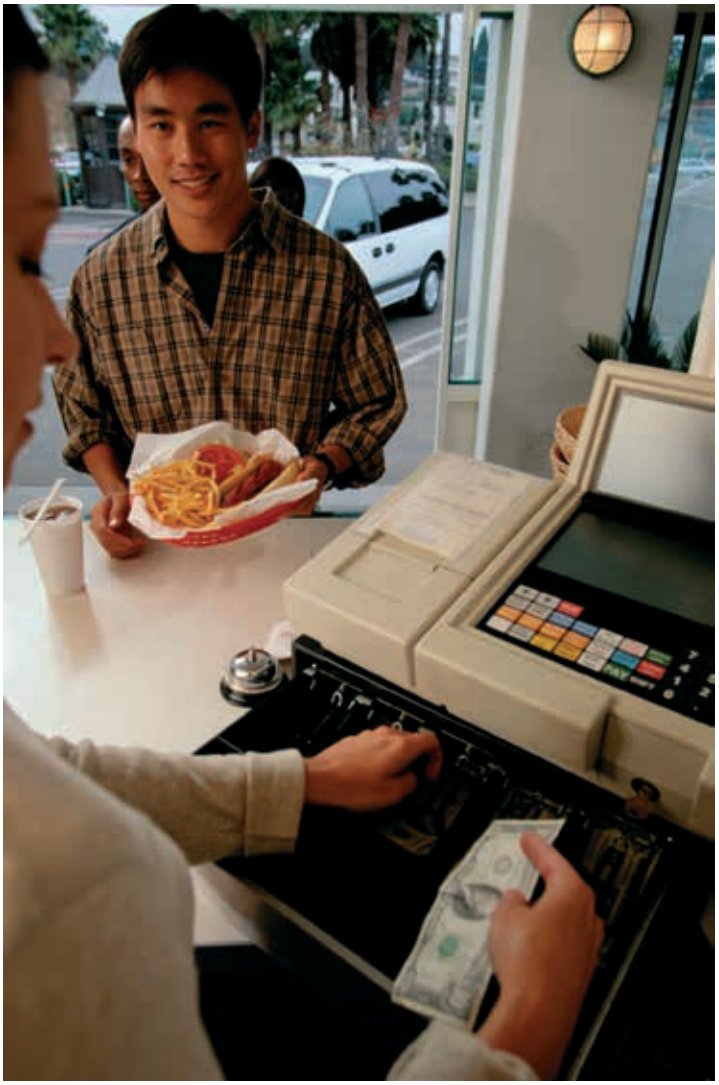


Fig 2.18 A point of sale terminal

A stock control system must keep an up-to-date record of all the stock held and place orders for fresh deliveries if stock runs low. Large shops, supermarkets and factories use computerised stock control systems.

Stock control is important as:

- Adequate stocks must be maintained to supply a customer with goods with minimum delay. If customers find goods are regularly out of stock they will go elsewhere.
- Goods must not be overstocked. By keeping stocks to a minimum, a business can limit the amount of money invested in stock and also reduce the risk of stock deteriorating before it can be sold. Minimum stock levels also reduce storage costs such as warehousing, heating, lighting and security.

A real-time stock control system

Input: The operator at a POS terminal only needs to pass the barcode on each item past a laser scanner. The scanner reads the code number stored in the barcode and sends it directly to a computer. The computer checks the code and, if it is valid, looks up the product's name and price in data files held on disk. The name and price are sent back to the POS terminal.

In this way the POS terminal can print out an itemised receipt. Note that with POS systems, prices are usually only marked on shelves and not on individual items, which can cause customer confusion. Each terminal has a keypad or keyboard that can be used if a barcode cannot be read.

Processing: As each item is sold, the stock files are updated so that customer service can be much quicker, reducing queues. Few mistakes can be made in charging customers. Prices can be changed easily. A fully itemised bill can be provided for the customer. No staff are needed for counting stock on shelves.

Output: Orders are printed when stock levels reach a re-order point. Customer receipts are printed using thermal printers.

Questions

1. Give an example of the processing required to determine if a customer's bill is overdue.
2. Give three examples of data that sensors can be used to measure.

Multiple choice questions

1 The manipulation of data to obtain information is called:
- **a** input
- **b** output
- **c** processing
- **d** storage.

2 Information that represents a 'whole' picture of a problem or solution is:
- **a** accurate
- **b** complete
- **c** relevant
- **d** timely.

3 A _____ check compares the contents of two or more fields to make sure that they make sense.
- **a** consistency
- **b** range
- **c** data type
- **d** format.

4 Hardware errors can be caused by any of the following, *except*:
- **a** bad sectors on a hard disk
- **b** corrupt RAM
- **c** power surge
- **d** program malfunction.

5 Which of the following checks is used to detect doubtful data?
- **a** length
- **b** range
- **c** data type
- **d** reasonableness.

6 An area code must contain three digits only. The following checks are suitable, *except*:
- **a** data type
- **b** format
- **c** length
- **d** range.

7 A _____ file is a temporary file which is used to _____data in the main file.
- **i** transaction, delete
- **ii** transaction, update
- **iii** master, delete
- **iv** master, update.

8 To find a record in a file using a sequential search, repeatedly _____ a record until the required record is found.
- **a** read
- **b** save
- **c** write
- **d** update.

9 A fully indexed file contains the record key and the storage address of the:
- **a** average key value
- **b** exact key value
- **c** highest key value
- **d** lowest key value.

10 The access method that calculates the exact address (location) of a record is *most* suitable for which type of file structure?
- **a** serial
- **b** sequential
- **c** fully indexed
- **d** direct access.

Short answer questions

11 A data logging system is used to record, at specific intervals, the temperature of the water in an aquarium.
- **a** State one item of hardware that can be used to capture the temperature readings.
- **b** Explain what the logging system could output if the temperature of the water is too high.
- **c** Describe one advantage of monitoring the water using this data logging system.

12 Aaron uses a login screen as shown below:

- **a** Give two types of applications that would require a user to log on.
- **b** Describe two security measures used on the screen displayed.
- **c** Explain why the password must be entered twice.
- **d** Explain whether re-entering the password twice is an example of verification or validation.
- **e** Explain two messages that the system could show based on the information entered by the user.
- **f** After Aaron has entered his username and password, he needs to press one of the two buttons on the screen to continue. What type of screen is Aaron using?

13 A student enters the following data from an experiment on a sheet that contains the headings and instructions on what to enter:

Results sheet

Experiment number	Start temperature	End temperature	Temperature difference
1	40	50	110
b	50	65	15
3	52	68	Yes

- **a** Three errors were made in the results sheet. Explain these errors.
- **b** Explain each of the following checks as it relates to the results sheet above:
  - **i** data type check
  - **ii** range check.
- **c** Describe how the student could confirm that the data from the table is equivalent to the data on the sheet.
- **d** The four statements below show the different descriptions during the experiment when completing the sheet. Match each of the descriptions with the most suitable term from the list below:

  Terms: source document/turnaround document/ machine-readable document/soft copy/hard copy

  Statements:
  - **i** The blank sheet before the student writes in results from the experiment.
  - **ii** The sheet is scanned and saved as a PDF file.
  - **iii** Results from the data on the sheet are used to produce a set of charts which are viewed on a monitor.
  - **iv** The completed sheet and charts are printed.
- **e** A teacher needs to print 70 copies of the sheet for the next class.
  - **i** Explain which type of printer would be suitable for this task.
  - **ii** No pages are being printed, even if the teacher selects the print icon repeatedly. Explain two possible causes of this problem.

14 Consider the following illustration of a supermarket, where items are placed in aisles:

Aisle	Locator	Item #
1	35	25 – Carrots 27 – Cucumber 29 – Lettuce 32 – Sweet pepper
2	50	38 – Disinfectant 40 – Liquid soap 43 – Scrub buds 47 – Sponges

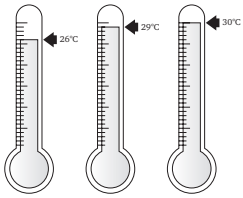
- **a** State the path taken to locate the:
  - **i** lettuce
  - **ii** sponges.
- **b** Suppose you were looking for item #30 (ginger).
  - **i** State the aisle and locator that you would choose and the item # that would indicate whether it is there or not.
  - **ii** What does the number of the locator represent?
- **c** Explain whether this is an example of direct, sequential or index-sequential file organisation.

Reinforce learning with the Workbook

2 *Information processing*

Data and information

1. Consider the following comparison of readings.



   a. What is the general name of the three objects in the figure? .....
   b. Explain the difference between data and information. .....
   .....
   c. Write two examples of data. .....
   d. Describe a source of data shown in the figure. .....
   e. Write a statement of information based on the figure. .....
   .....

2. The object shown here is used to represent information about a product. This barcode data is read and the information is recorded and interpreted by a device each time a product is purchased. The manager receives a weekly report on the sales of the products from the computer system.

   a. State the name of the object. .....
   b. Give two examples of products that could have this object on them. .....
   .....
   c. Describe a hardware device used to extract the data from this object.
   .....
   d. Explain whether the object represents direct data input or manual data input at the point of sale. .....
   .....
   e. State two items of information the object may store.
   .....
   f. State two items of information that the weekly sales report may contain.
   .....

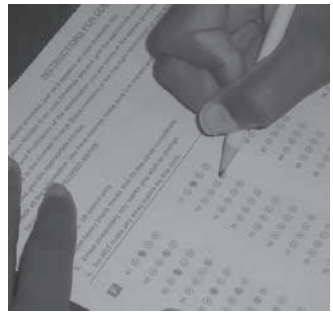
What's on the online support?

- Answers to the End of Chapter Exam-style questions
- Additional Exam-style questions with sample answers and examiner analysis
- Interactive Paper 1 multiple choice test, Practice Paper 2, worked SBA Paper 3, with advice on online marking, SBA alternative Paper 3
- Answers to Workbook questions
- Comprehensive glossary of terms

Additional practice in the Workbook extends the skills students have learnt in class.

Methods of data capture

1. A student shades certain areas on a sheet to represent his responses to a set of questions.



   a. Indicate which of the following terms matches the descriptions:

human readable document	turnaround document	source document
machine readable document		

   i. A sheet that has not yet been shaded is an example of a
   .....
   ii. The sheet after the student has marked his responses is an example of a
   .....
   iii. The sheet with responses marked, submitted to the machine, is an example of a
   .....
   iv. The machine prints the numbers as a table and plots them on a chart. These results are examples of a
   .....
   b. State the input device that is used with this type of sheet.
   c. Describe how the student could confirm that the data from the printout is equivalent to the data on the sheet. .....
   .....
   d. Describe an error that the student could have made when shading the answers on the sheet.
   .....

Sample pages from IT for CSEC© WorkBook, Unit 2

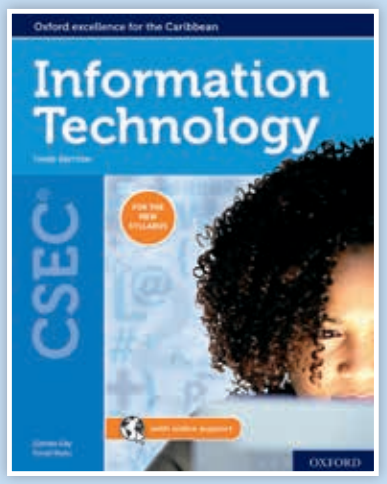
View a copy of the book or place your order

Contact your local bookstore:

Available in the series

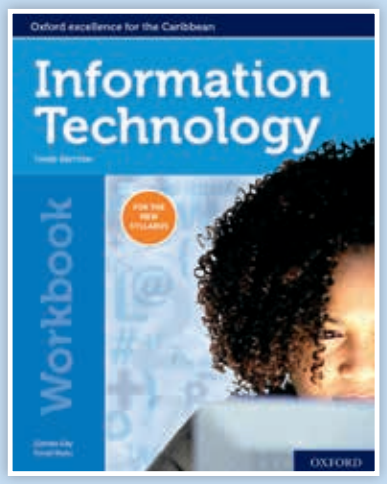
Information Technology CSEC® - Third Edition

Student Book and Study Guide also available as eBooks

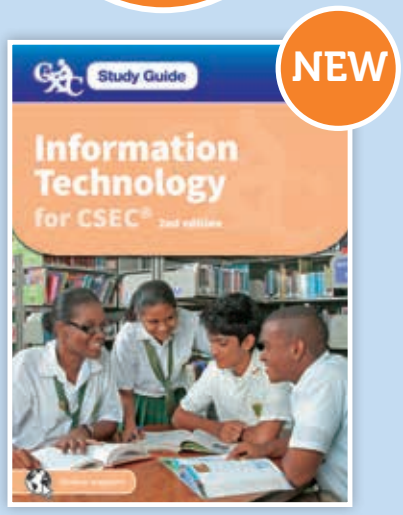


Student Book
978 019 843716 1

Student Book eBook
978 019 843719 2



Workbook
978 019 843720 8



Study Guide
978 019 843721 5

Study Guide eBook
978 019 843724 6

K45260