Diploma Programme
Programme du diplôme
Programa del Diploma

Computer science
Higher level
Paper 2

5 May 2025

Zone A morning | **Zone B** morning | **Zone C** morning

1 hour 20 minutes

Instructions to candidates

- Do not open this examination paper until instructed to do so.
- Answer all of the questions from one of the options.
- The maximum mark for this examination paper is **[65 marks]**.

Option	Questions
Option A — Databases	1 – 4
Option B — Modelling and simulation	5 – 8
Option C — Web science	9 – 13
Option D — Object-oriented programming	14 – 17

23 pages

Option A — Databases

1. Alpha Hospital is situated in a large city and has over 1000 staff. It stores its data in a relational database.

Some of the tables in this relational database contain data about the hospital staff, their roles, and the hospital departments.

Staff roles include doctor, nurse, pharmacist, radiologist, and support staff. Staff can only hold one role.

Departments include Accident and Emergency, Critical Care, Medical, General Surgery, Orthopaedics, and Ophthalmology. Staff can work in several departments.

The ROLE table, STAFF table, and DEPARTMENT table are shown in **Figure 1**.

Figure 1: Diagrammatic relationship for the ROLE table, STAFF table and DEPARTMENT table

ROLE
RoleID
PayGrade

STAFF
StaffID
RoleID
DepartmentID
FirstName
Surname
DateOfBirth

DEPARTMENT
DepartmentID
ManagerID
Building

(a) (i) State the primary key in the STAFF table. [1]

(ii) State a foreign key in the STAFF table. [1]

(b) Describe the relationships between the three tables in **Figure 1**. [2]

(c) Outline what a query is used for in a database. [2]

(d) Identify the steps to create a query to list the staff with the surname Waters who are on pay grade 17. The query must display **only** FirstName, Surname, and PayGrade. [4]

(e) Outline why an integer is an appropriate data type for the PayGrade field. [2]

(f) Explain **two** ways in which the database administrator can ensure the privacy of the hospital's staff data. [6]

(Option A continues on the following page)

(Option A continued)

2. Database design is a complex process that takes place in a range of phases. Different phases of database design use different schema.

(a) Describe the difference between a conceptual schema and a logical schema. [2]

(b) Explain the importance of a data definition language in implementing a data model. [2]

(c) Explain why data modelling is used during the development of a database. [4]

(d) Explain why both data validation and data verification are required to ensure the correctness of the data within a database. [3]

(e) Outline how data integrity is maintained during a database transaction. [2]

(f) Outline the role of relational integrity in maintaining data consistency within a database. [2]

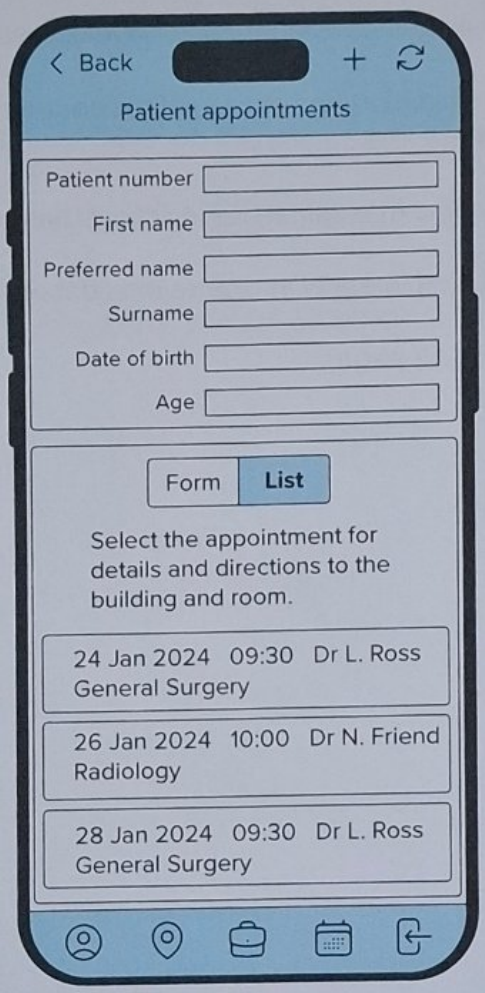
(Option A continues on the following page)

(Option A continued)

3. Alpha Hospital uses a database application to book appointments. Information can be shown in different formats.

Figure 2 shows an example of a patient's appointments displayed in the application.

Figure 2: A patient's appointments



The Age field is a derived field.

(a) (i) Outline **one** reason why a derived field would be used in a database. [2]

(ii) Describe how the derived field Age would be calculated. [2]

Patient information can be represented in the following format:

`PATIENT (PatientID, FirstName, Surname, PreferredName, DateOfBirth)`

(b) Outline the difference between first normal form (1NF) and second normal form (2NF). [2]

(c) Construct a database in third normal form (3NF) for all of the data shown in **Figure 2**. You should use database notation as shown in the PATIENT table. [6]

(Option A continues on the following page)

(Option A continued)

4. Alpha Hospital is part of a wider health authority that manages several hospitals in the region. This health authority is planning to integrate the databases held by each hospital into a single database.

The health authority is considering a range of database models for this integration. One model is a relational database.

(a) Identify **two other** database models that could be used to integrate the databases held by each hospital into a single database. [2]

The National Health Authority wants to coordinate all data records in a data warehouse. Each regional health authority will provide the data that will be mined.

(b) Define the term *data warehouse*. [2]

(c) Explain why the extract, transform, load (ETL) process is used to prepare data for data warehousing. [4]

(d) Outline the importance of timestamping in a data warehouse. [2]

(e) Describe how data in a data warehouse is updated in real time. [2]

(f) Describe the process of deviation detection. [2]

Data mining techniques are being used in the data warehouse to assist the detection of patterns in the health data. The pattern detection approaches being used include cluster analysis and association rules.

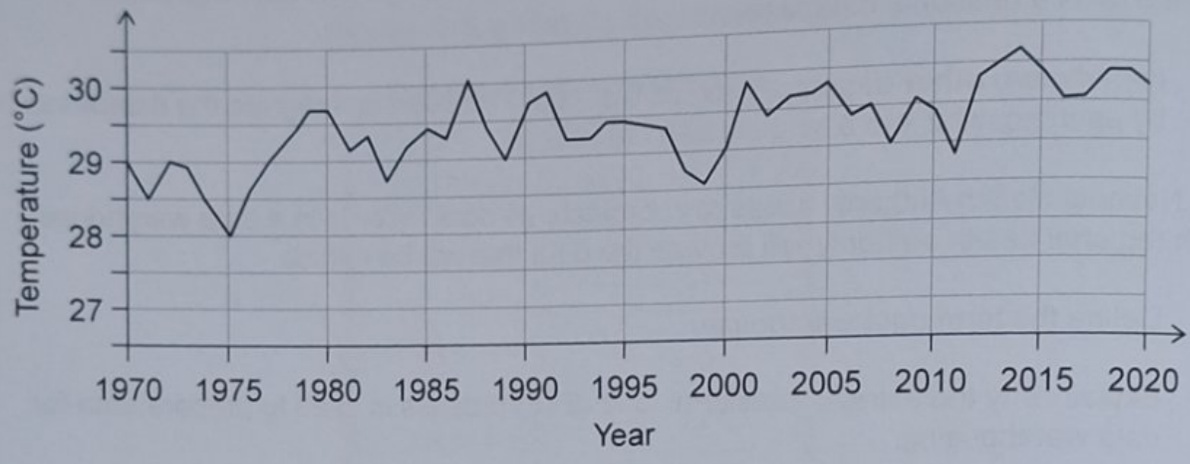
(g) Compare and contrast cluster analysis and association rules as methods of identifying patterns in data mining. [6]

End of Option A

Option B — Modelling and simulation

5. Global warming can be measured over time using mean temperatures. **Figure 3** shows the mean daily maximum temperatures from 1970 to 2020 for Nauru, an island in the Pacific Ocean.

Figure 3: Mean daily maximum temperatures for Nauru, 1970–2020



(a) State **two** variables that are necessary for the mean daily maximum temperature to be calculated. [1]

(b) Identify the steps that would be used to create the diagram exactly as shown in **Figure 3** using the data in a spreadsheet. [3]

(c) Identify **two** reasons why the mean daily maximum temperature data is presented in graphical form. [2]

An automated system is used to collect the temperature data for Nauru once an hour for one year.

(d) Outline **one** reason why the temperature data is collected once an hour rather than at shorter intervals, such as once a minute. [2]

(e) Describe the steps that should be used to store the data collected for one day (24 hours) in suitable parallel one-dimensional (1D) arrays.

The time the data was collected must be easy to identify. **Do not** write a pseudocode algorithm. [5]

(Option B continues on the following page)

(Option B, question 5 continued)

The temperature data is downloaded every day and collated into a master file.

The data from the master file is loaded into a suitable array for that 24-hour period.
The following statistics are calculated:

- Maximum temperature
- Minimum temperature
- Mean temperature

As Nauru is very close to the equator, the length of the day changes very little throughout the year. For the purposes of part (f), the lengths of its day and night are:

- Day: 07:00 to 18:59 inclusive (12 hours)
- Night: 19:00 to 06:59 inclusive (12 hours)

(f) Construct a pseudocode algorithm to calculate the:

- maximum temperature
- minimum temperature
- mean temperature
- mean night-time temperature.

Assume the arrays for the time of day of the reading and hourly temperature readings have already been set up and populated as parallel 1D arrays. [8]

(Option B continues on the following page)

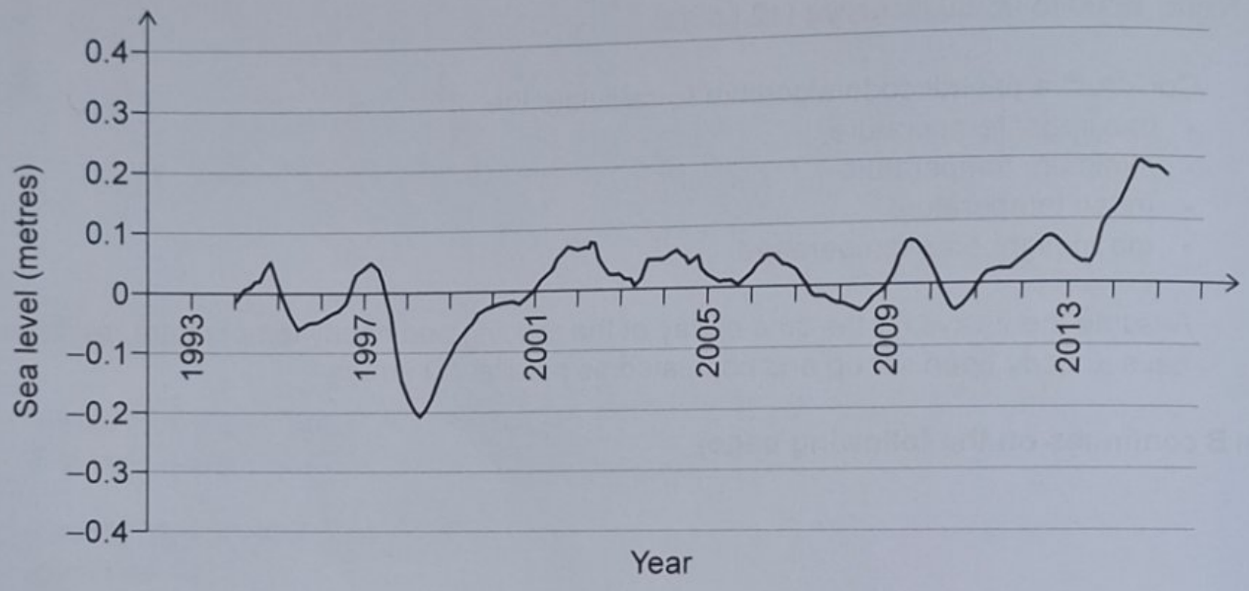
(Option B continued)

6. The global increase in mean temperatures is causing concern, and governments are using computer models to determine potential future changes.

Many islands in the Pacific Ocean are close to or below sea level and have observed an increased incidence of coastal flooding. This suggests there is a relationship between the increase in mean temperatures and the increase in mean sea levels.

Figure 4 shows the mean sea level of Nauru from 1994 to 2015.

Figure 4: Mean sea level of Nauru, 1994–2015



If the mean sea level rises, the probability of coastal flooding will also increase. This will put many of the inhabitants of Nauru at risk.

It has been proposed that a simulation be developed to show the effects of rising temperatures on the extent and frequency of coastal flooding.

(a) Distinguish between a model and a simulation. [2]

(b) Describe how to identify the rules required to create a simulation from the mean temperature and mean sea level data. [3]

(c) Evaluate how test cases could be used to effectively validate the accuracy of this proposed simulation. [6]

(d) Discuss the advantages **and** disadvantages of using a simulation for decision making in the coastal areas of islands in the Pacific Ocean. [5]

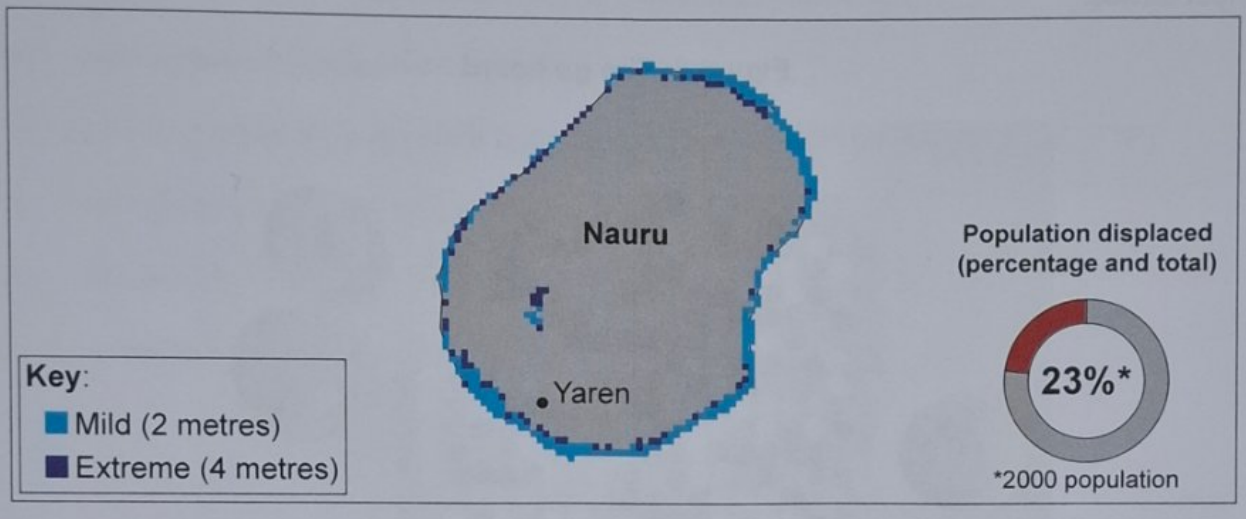
(Option B continues on the following page)

(Option B continued)

7. Organizations such as Earth.Org have raised concerns about the rate of sea level increase for Nauru. They stated, "Sea level rise is 2 to 3 times faster around Nauru than the global average, putting its freshwater supplies and crops at risk of saltwater contamination. Already reliant on economic aid, Nauru's basic resource needs may have to be acquired externally for life to be sustained on the island."

The 2D visualisation in **Figure 5** shows the projected impact of mild and extreme sea level increases on Nauru by 2100.

Figure 5: Projected impact of mild and extreme sea level increases on Nauru by 2100



(a) Define the term *visualization*. [1]

There are proposals to develop a 3D visualization of the impact of rising sea levels on Nauru.

(b) Outline the relationship between images stored in memory and 3D visualizations. [2]

(c) Discuss the time and memory considerations of 3D animation in the proposed 3D visualization for Nauru. [5]

(Option B continues on the following page)

(Option B continued)

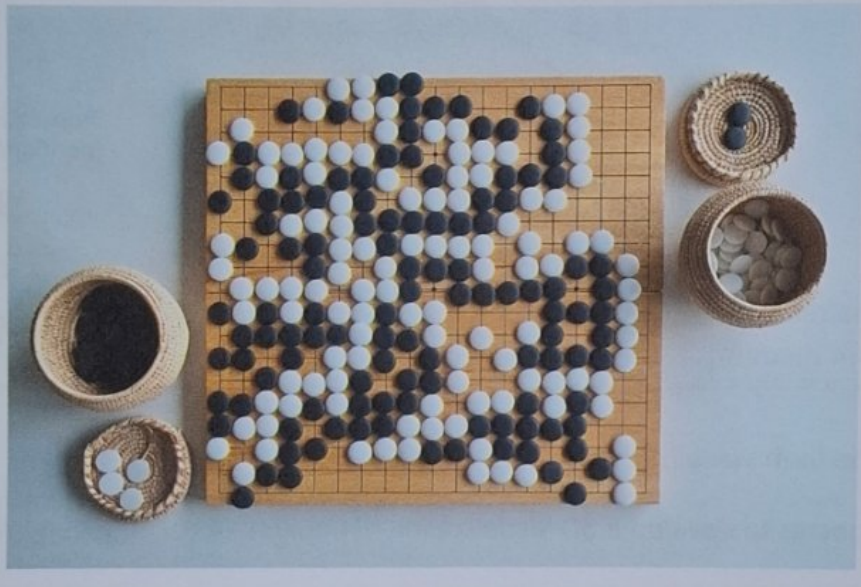
8. Genetic algorithms are based on Darwin's theory of natural selection. The process selects the fittest individuals.

(a) Identify **three** of the phases in a genetic algorithm. [3]

Developments in machine learning and neural networks have led to simulations that are able to beat the leading human players of games. For example, Deep Blue beat Gary Kasparov in chess, and Alpha Go beat Fan Hui in go (a Chinese strategy board game).

The neural network used to defeat Fan Hui used the go board (see **Figure 6**) as an input device.

Figure 6: The go board



Within the neural network there is a policy layer that selects the next move. There is also a value network that predicts the winner of the game.

(b) State **two** components of a neural network. [2]

(c) Identify the steps that could be used to train the neural network used for the simulation of go to recognize patterns of play. [5]

(d) Explain how supervised learning and unsupervised learning could lead to different outputs from the neural network. [5]

Neural networks can be used in a variety of contexts, such as for predicting outcomes of board games like go and chess or for natural language processing.

(e) Explain how advances in natural language processing have improved the accuracy of the predictions of neural networks. [5]

End of Option B

Option C — Web science

9. Alexia and Jay are discussing their use of online resources to prepare for their IB examinations. They often visit websites such as BBC Bitesize.

Alexia refers to accessing the websites as "surfing the internet" and Jay refers to the resources as being "on the web".

(a) Distinguish between the internet and the World Wide Web. [2]

The uniform resource locator (URL) for the BBC Bitesize website is as follows:

https://www.bbc.co.uk/bytesize/index.htm

(b) Identify **two** characteristics of a URL. [2]

Jay wants to transfer a file to Alexia and suggests they use file transfer protocol (FTP).

(c) Identify **two** characteristics of file transfer protocol. [2]

(d) Explain how a web browser functions. [3]

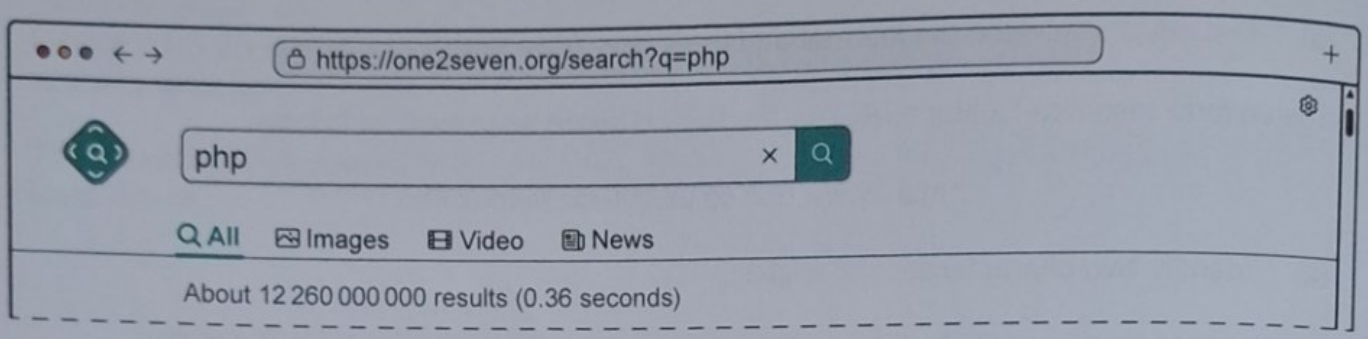
(Option C continues on the following page)

(Option C continued)

10. Alexia and Jay are researching the web development language PHP.

They type the phrase "PHP" directly into a web browser (see **Figure 7**).

Figure 7: "PHP" typed into a web browser and redirected to a search engine

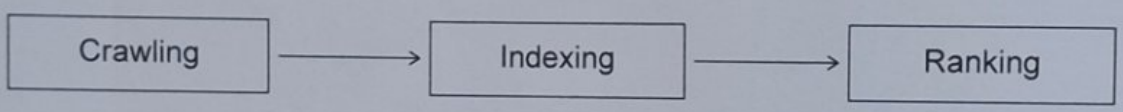


The web browser redirects them to a popular search engine, which executes a search.

(a) Define the term *search engine*. [1]

The operation of a search engine can be divided into three steps (see **Figure 8**).

Figure 8: The three steps in the operation of a search engine



(b) Describe how a web crawler functions. [2]

(c) Outline why keywords are important for web indexing. [2]

(d) Discuss whether an organization should use black hat search engine optimization (SEO) techniques to improve the ranking of its website. [6]

Search engines return a very large number of results, but many of the web pages are not useful. The search needs to be refined.

Alexia and Jay's teacher recommended that they use an online database that accesses the deep web.

(e) Distinguish between the surface web and the deep web. [2]

(Option C continues on the following page)

(Option C, question 10 continued)

One of the online databases provides Alexia and Jay with the following code:

```
<?php
    if(isset($_FILES['CV'])){
        $errors= array();
        $file_name = $_FILES['CV']['name'];
        $file_size =$_FILES['CV']['size'];
        $file_tmp =$_FILES['CV']['tmp_name'];
        $file_type=$_FILES['CV']['type'];

        $file_ext=strtolower(end(explode('.',$_FILES['CV']['name'])));

        $extensions= array("pdf","doc","docx");

        if(in_array($file_ext,$extensions)=== false){
            $errors[]="This file extension not allowed";
        }

        if($file_size > 2097152){
            $errors[]="File size must be under 2 Mb";
        }

        if(empty($errors)==true){
            move_uploaded_file($file_tmp,"CV/".$file_name);
            echo "Success";
        }else{
            print_r($errors);
        }
    }
?>
<html>
<body>
    <h1>Curriculum Vitae</h1>
    <form action="" method="POST" enctype="multipart/form-data">
        <input type="file" name="CV" />
        <input type="submit"/>
    </form>
</body>
</html>
```

(f) Identify **four** steps that take place during the processing of this PHP code. [4]

The PHP code is processed on the server.

(g) Explain why an organization would choose to use server-side processing rather than client-side processing when delivering content to the client. [3]

(Option C continues on the following page)

(Option C continued)

11. A social media file-sharing website allows users to upload their original video content.

The site provides an upload module that uses lossless compression.

(a) Outline how lossless compression maintains the quality of a media file. [2]

When a file is decompressed, the content is rendered to fit a standard set of sizes, aspect ratios, and frame rates. Standard video formats, such as MP4, are used for the output of media.

The file format MP4 or MPEG-4 is an open standard for media files.

(b) Identify **two** characteristics of an open standard. [2]

Social media sites use a distributed system where the data is stored in many countries.

In the European Union, Argentina, and the Philippines the right to be forgotten has been established as law. The right to be forgotten means a person has the right to have private information removed from databases and applications so it cannot be found in an internet search.

(c) Discuss the impact of the decentralized web on an individual's right to privacy. [6]

Xero, a small software company, was established in 2006 in Wellington, New Zealand. It produced an easy-to-use accountancy service for small to medium enterprises. This product was based on the software-as-a-service (SaaS) model and is sold in a subscription format all over the world. Customer data is securely stored online as part of the subscription cost.

(d) Explain how developments in the web have enabled small companies such as Xero to have a global reach. [6]

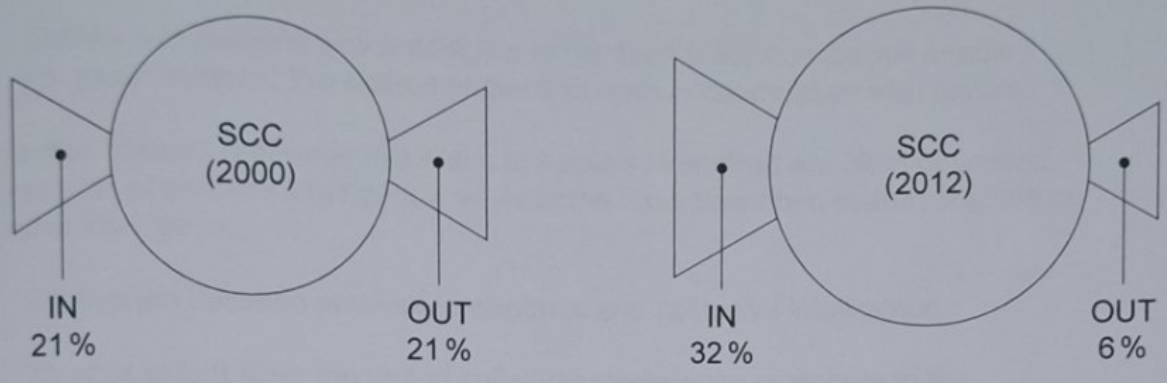
(Option C continues on the following page)

(Option C continued)

12. The bowtie structure of the web is described as a web graph.

Figure 9 shows the bowtie structure of the web based on two web crawls – one in 2000 and one in 2012.

Figure 9: Bowtie structure of the web, 2000 and 2012



(a) (i) Define the term *node*. [1]

(ii) Define the term *edge*. [1]

(b) Define the term *strongly connected core (SCC)*. [1]

Compare the 2000 and 2012 structures in **Figure 9**. The diameter (relative importance) of the SCC has changed.

In 1980, Robert Metcalfe proposed that the influence or effect of a telecommunications network is influenced by the number of connections in that network.

This can be expressed as a mathematical equation:

$$\text{effect} = \frac{n(n-1)}{2}$$

Where:
- n is the number of nodes
- effect is the number of edges.

(c) With reference to the bowtie structure of the web, explain whether Metcalfe's law is a useful measure in determining the connectedness of the web. [3]

(Option C continues on page 17)

(Option C continued)

13. The evolution of the web has been described as taking place in three stages: its beginnings as Web 1.0, the static web; Web 2.0, the "social or read-write" web; and finally Web 3.0, the Semantic web.

(a) Describe the aims of the Semantic Web. [2]

Ontologies are a technology associated with the Semantic Web.

(b) Outline **two** reasons why ontologies rather than folksonomies will enable computer systems like search engines to work in cooperation with people. [4]

Connected devices like Amazon's Alexa or Apple's HomePod are often described as examples of ambient intelligence, while some have described search engines as collective intelligence.

(c) Distinguish between ambient intelligence and collective intelligence. [2]

(d) To what extent does the use of collective intelligence contribute to the development and evolution of search engines? [6]

End of Option C

Option D — Object-oriented programming

14. A company sells healthcare products. Each product type is associated with a particular brand and brand price.

A system is designed to manage product sales, customers, and suppliers.

The `Product` class keeps details of a product. The following shows part of the code for this class:

```
public class Product {

    private String prodCode; // eg X123
    Private String prodType; // eg Sunscreen
    private String prodDescription; // about the product
    private Brand prodBrand; // an object of type Brand
    private int prodSale; // number of units sold

    // Constructor
    // code missing for constructor method

    public int getProdSale(){
        return prodSale;
    }

    public Brand getProdBrand(){
        return prodBrand;
    }

    // all accessor and mutator methods are present but not shown

} // end of Product class
```

The `Brand` class keeps details of a particular brand. The following shows part of the code for this class:

```
public class Brand {

    private String brandName; // eg Safesun
    private float brandPrice; // price of the product of this brand

    public Brand(String brandName, float brandPrice) {
        this.brandName = brandName;
        this.brandPrice = brandPrice;
    }

    public float getBrandPrice() {
        return brandPrice;
    }

    // all accessor and mutator methods are present but not shown

} // end of Brand class
```

(Option D continues on the following page)

(Option D, question 14 continued)

(a) (i) Define the term *private*. [1]

(ii) Define the term *accessor method*. [1]

(iii) Construct an accessor method in the `Product` class that returns the description for a product. [2]

(b) (i) Construct a UML diagram for the given `Product` class. [3]

(ii) Construct the code for the constructor method for the `Product` class that initializes all attributes for a new product. [3]

(c) (i) Construct the code to create an instance of the `Brand` class that has the brand name Safesun and a brand price of 2.17. [3]

(ii) Describe the difference between a class and an instance of a class. [2]

(d) Identify **two** features of modern programming languages. [2]

15. (a) (i) Define the term *primitive data type*. [1]

(ii) Outline **one** advantage of using a primitive data type, such as `int`. [2]

The `ProductManagement` class has the main method and other methods to generate the information required:

```
public class ProductManagement {

    private Product[] allProducts = new Product[25];

    public void sortProducts() // sort in descending order of prodSale {
        // code missing
    }

} // end of ProductManagement class
```

(b) Construct code for the method `sortProducts()` to sort the `allProducts[]` array in descending order of `prodSale`.

You must make use of the selection sort algorithm. [6]

(Option D continues on the following page)

(Option D continued)

16. (a) (i) Outline **one** advantage of polymorphism. [2]

(ii) Outline **one** advantage of encapsulation. [2]

(iii) Outline **one** disadvantage of inheritance. [2]

An invoice is created every time a customer purchases one or more products.

The `Invoice` class keeps details of each invoice. The following shows part of the code for this class:

```
public class Invoice {

    private String invoiceID; // identifies a unique invoice
    private static Product[] products = new Product[20]; // list of
products purchased
    private static int[] prodQuantity = new int[20]; // number of items
of a particular product purchased
    private boolean qualifiesForDiscount; // default value is false
    private int numOfProducts; // how many products in this invoice

    // constructor is defined, code not shown

    public String getInvoiceID(){
        return invoiceID;
    }

    // all accessor and mutator methods are present but not shown

    public void addProduct(Product product, int quantity) {
        // code missing
    }

    public void setQualifiesForDiscount(){
        // if total value of purchases is more than 3000,
        // qualifiesForDiscount value is set to true
        // code missing
    }

} // end of Invoice class
```

(b) Describe how encapsulation has been used in this code. [2]

(Option D continues on the following page)

(Option D, question 16 continued)

(c) Construct the method `addProduct(Product product, int quantity)` that will update an invoice.

The method should:
- Update the `products` array
- Update the `prodQuantity` array
- Increment the `numOfProducts`. [3]

If the total value of the purchases is greater than 3000, the invoice qualifies for a discount.

(d) Construct code for the method `setQualifiesForDiscount()` to change the status of an invoice.

The method should:
- calculate the total value of an invoice
- change the value of `qualifiesForDiscount` if needed. [6]

(e) Outline **one** advantage of using modularity in program development. [2]

(Option D continues on the following page)

(Option D continued)

17. The `Supplier` class gives the details of suppliers. The following shows part of the code for this class:

```
import java.util.LinkedList;

public class Supplier {
    private String supplierName;
    private String supplierCountry;
    private String[] productNames = new String[10];

    public Supplier(String supplierName, String supplierCountry,
String[] productNames) {
        this.supplierName = supplierName;
        this.supplierCountry = supplierCountry;
        this.productNames = productNames;
    }

    public String getSupplierName(){
        return supplierName;
    }

    public String getSupplierCountry(){
        return supplierCountry;
    }

    // all accessor and mutator methods are present but not shown

    public String displayData() {
        System.out.println("Supplier: " + supplierName + ", Country: "
+ supplierCountry);
    }

} // end of Supplier class
```

The `SupplierManager` class has methods to generate the information required. The following shows part of the code for this class:

```
public class SupplierManager {
    LinkedList<Supplier> supplierList;

    public SupplierManager() {
        supplierList = new LinkedList<>();
    }

    public void addSupplier(Supplier newSupplier) {
        // code missing
    }

    public void displayList() {
        // code missing
    }

    public static int countOfSuppliers(LinkedList<Supplier>
supplierList, String country, int n) {
        // code missing
    }

} // end of SupplierManager class
```

(Option D continues on the following page)

(Option D, question 17 continued)

a) (i) Outline **one** advantage of using library collections when developing a program. [2]

(ii) Construct the method `displayList()` that uses the `displayData()` method and collections method to output the details of all the suppliers in the linked list. [2]

A linked list, `supplierList`, is used to store `Supplier` objects.

(b) Construct the method `addSupplier(Supplier newSupplier)` that will add new suppliers to a linked list ordered by supplier name. [3]

(c) (i) Define the term *recursion*. [2]

Many suppliers come from different countries.

A method exists that gives the size, `n`, of the supplier linked list.

(ii) Construct the recursive method `countOfSuppliers (LinkedList<Supplier> supplierList, String country, int n)` that returns how many suppliers are from this country. [5]

(iii) Outline **two** reasons why a recursive algorithm is not the most appropriate algorithm for the scenario described in part (ii). [4]

The programmer has made use of naming conventions throughout the code.

(d) Outline **one** advantage of using naming conventions. [2]

End of Option D