

Computer science
Higher level
Paper 2

Monday 8 May 2017 (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 – 12
Option D — Object-oriented programming	13 – 16

Blank page

Option A — Databases

- 1. A school stores details about the parents of its students in a database management system.
 - (a) Outline **two** reasons for choosing to use a database for this storage. [2]
 - (b) Identify **one** item of data that would be stored in this database management system. [1]
 - (c) Outline a technique that could be used to reduce data input errors. [2]

One of the functions of the system is to allow parents to pay for their children’s school lunches online. Parents are notified, by email, with the amount that they owe. After payment, a confirmation is sent to the parent by email.

- (d) Describe the interaction between the parent and the database management system when processing a payment. [5]
 - (e) Explain how the database is accessed when many parents pay online at the same time. [3]
- 2. (a) Define the term *relational database*. [1]
 - (b) Explain the role of entity-relationship diagrams when modelling data. [4]

TABLE_1 and TABLE_2, shown below, are taken from a school’s relational database. TABLE_1 holds data about all of the students at the school and TABLE_2 holds data about participants in the school chess club. The tables’ primary keys are underlined.

TABLE_1

<u>Student_ID</u>	EmailAddress
12	fred@mail.com
34	ng@mail.com

TABLE_2

<u>Member_ID</u>	Student_ID	Position
1	12	Captain
13	34	Secretary
11	21	Member

- (c) Limited to the data shown, outline an issue that may arise when using the tables. [2]
- (d) (i) Describe an issue that could occur after deleting, in TABLE_1, the record for the student whose Student_ID is 12. [2]
- (ii) With reference to the situation in part (d)(i), outline which property of data ensures that, in a relational database, the deletion of records has no unwanted effects. [3]

(Option A continues on the following page)

(Option A continued)

3. A phone company stores the SMS text messages that are sent between its own subscribers and only these subscribers.
- Each subscriber that sends messages has an associated data structure where details of their SMS text messages are stored.
 - The sender’s phone number and name, followed by the details of each individual SMS text message are collected as they occur.
 - The details stored for each SMS text message are the receiver’s phone number, name and text message.

For example, consider the following two data structures:

1111, fred, 1234, smith, Hi there, 1234, smith, Call me, ...
1234, smith, 1111, fred, Later, ...

The first data structure is associated to a sender whose phone number is 1111, and has sent two SMS text messages to another subscriber whose phone number is 1234.

The second data structure is associated to a sender whose phone number is 1234, and currently contains the details of just one message.

- (a) With reference to the data organization in the two examples,
- (i) outline **one** limitation of using a table to store this collection of data exactly as it is illustrated. [2]
 - (ii) construct a **single** table that could store all the data for each message separately, for each subscriber. Your answer should include the example data shown above. [4]
- (b) Define the term *data redundancy*. [2]
- (c) Construct **two** tables to remove any data redundancy from part (a)(ii). [4]
- (d) Using the tables from part (c), suggest the steps for a query to obtain, for any SMS text message, the name of the receiver. [3]

The phone company hires a database administrator.

- (e) Outline **one** function that a database administrator should perform regularly. [2]
- (f) Discuss an ethical issue that could arise if external agencies could obtain and analyse the phone company’s data. [3]

(Option A continues on the following page)

(Option A continued)

4. An educational authority, in a certain country, administers a school system that includes more than 1000 schools. Each school has stored data over the past 20 years relating to their current and former students, including their name, age, gender, date of birth, address, health records, and examination results and test scores. This data is to be added to the educational authority’s data warehouse.

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

The educational authority is interested in comparing how examination results and test scores in various subject areas have changed over time.

(b) Describe the process that would need to be followed to move the data from each school into the data warehouse. [4]

(c) Explain the process that could be used to locate schools that have low test scores in mathematics for a particular age group. [4]

(d) Define the term *data mining*. [2]

(e) Compare the use of cluster analysis and association analysis in attempting to understand the distribution of scores in grade 6 reading tests over the past 10 years. [4]

(f) Suggest **one** negative social impact that may arise from using cluster analysis and association analysis in this data mining if the results are published. [4]

End of Option A

Option B — Modelling and simulation

5. A local sports league arranges a series of games between four teams over one year. Each team plays every other team 10 times. Using the number of games won and lost in previous years, a table of probabilities is calculated and held in an array PROB. The contents of PROB is represented by the following diagram.

PROB		1	2	3	4
	1		0.5	0	0.5
	2	0.2		0.8	0.5
	3	0.9	0.1		0.4
	4	0.3	0.5	0.4	

The content of PROB[2, 3] is 0.8 and indicates that the probability of team 2 winning against team 3 is 0.8. For the 10 games played between these two teams it is likely that team 2 will win 8 of them.

- (a) With reference to the table, state the probability of team 3 winning against team 2. [1]

PROB[*i*, *j*] is the probability of team *i* winning against team *j*.

- (b) Using this notation, identify the probability of a draw between team *i* and team *j*. [2]

- (c) State the total number of games that each team plays in one year. [1]

Based on these probabilities, a table is set up in a spreadsheet to store the expected number of games that each team will win and lose this year.

The table, PROB10, is in the same format using rows 1–4 and columns A–D as shown below.

PROB10		A	B	C	D
	1		5	0	5
	2	2		8	5
	3	9	1		4
	4	3	5	4	

A macro is created, in this spreadsheet, to calculate and store the expected number of draws for each team, so that cell A1 holds the total number of times team 1 is expected to draw, B2 holds the total number of times team 2 is expected to draw and so on.

- (d) With the use of an illustration, demonstrate the processes that would be carried out by the macro. [3]

(Option B continues on the following page)

(Option B, question 5 continued)

During the year, results of each game are used to complete a separate table, PLAYED. For each team, PLAYED stores the number of games lost, won and drawn against each other team in a format similar to PROB10.

After each game, information including the team numbers and scores, is sent in the form i, j, m, n . For example 2, 3, 6, 1 indicates that team 2 scored 6 and team 3 scored 1.

- (e) Using the data i, j, m, n construct an algorithm to insert the correct data into the table PLAYED. You can assume that the table contains zero values at the beginning of the year. [4]

Each team is ranked based on how many games they have won that year. This is then compared with the data in the **expected** tables.

- (f) With reference to this scenario, distinguish between a mathematical model and a simulation. [4]
- (g) Discuss the benefits of comparing the expected results with the actual results. [4]

(Option B continues on the following page)

(Option B continued)

6. A reservoir is being planned, to provide water for a local town. A simulation is carried out to aid the planning process.

- The initial volume in the reservoir is preset to an appropriate level.
- The input will be from rivers and will vary according to the climate.
- The output will be through a large pipe that is large enough to carry enough water for the daily needs of the people in the town.

To avoid the reservoir exceeding a dangerous level, a series of outlets are planned to divert water away from the reservoir if the level of water reaches a certain limit.

The following variables, each calculated at the same moment in time, are used in the simulation:

- V represents the volume of water in the reservoir
- R represents the volume of input from the rivers and any rainfall
- T represents the volume of output via the pipe that sends water to the town
- E is the excess volume emptied through the planned outlets, if the volume in the reservoir exceeds D .

The volume that must not be exceeded is represented by D , and its value is fixed.

(a) State the relationship between V and D that avoids any danger. [1]

The simulation is used to determine the volume of water that needs to be emptied through the outlets when the danger level of the reservoir is reached.

(b) Describe the factors to consider when identifying suitable ranges of values of R , T and E to use in the simulation. [6]

The results of the simulation are to be displayed graphically showing the estimated water level.

(c) Discuss the way in which a series of graphs could be created over time, and interpreted. [6]

Software is used to design the reservoir incorporating all points at which water can enter and exit the reservoir and all materials to be used.

(d) Explain how the results of the simulation could be used to help the design of the reservoir. [4]

(Option B continues on the following page)

(Option B continued)

7. (a) Outline a situation where 2D visualization would be more appropriate than 3D visualization. [2]

An artist is planning to create a statue, but decides to do some planning before she starts.

- (b) Outline the advantages, to the artist, of using a computer aided design (CAD) and graphic processing software for 3D modelling, instead of a series of sketches drawn by hand, to experiment with designs. [3]
- (c) Outline the basic steps used in the process of rendering to give a 3D visualization of the final proposed statue. [4]
8. A researcher uses online methods, including emails, to get opinions from academics on the likely future advances in technology. Over 600 000 posts are subjected to text analysis using natural language processing (NLP) in order to summarize the opinions.

- (a) Outline **two** advantages of using NLP techniques in this context. [4]
- (b) Identify **two** problems that need to be resolved when posts from new sources are input for analysis. [2]

Supervised learning is used to train a neural network to analyse the data.

- (c) Suggest, for this scenario, a method of developing a fitness function before submitting all of the data to NLP. [2]
- (d) With the aid of a diagram, explain how this fitness function can be used with a neural network to analyse the complete data. [6]
- (e) Discuss the reasons why unsupervised learning may have been a better choice to train the neural network in this situation. [6]

End of Option B

Option C — Web science

9. An e-business, X-Sprintz, sells running shoes online. The web pages of their website contain several elements.

- (a) Outline **two** reasons why X-Sprintz would put their banner across the top of their web pages. [4]
- (b) Identify **one** piece of information that might be present in the footer of X-Sprintz’s web pages. [1]

A user is searching the website for a product. To narrow the search among the many products, the user can interact with the website by entering information into a web form that has the following code.

```
<html><body>
<h3>Search Form</h3>
<form action="process.php" method="post">
  <select name="item">
    <option>Spikes</option>
    <option>Trail</option>
    <option>Endurance</option>
  </select>
  <p> My size is:
  <input name="size" type="text" />
  <input type="submit" value="Send it"/></p>
</form>
</body></html>
```

- (c) (i) Identify the language used to write this code. [1]
- (ii) Describe the sequence of actions carried out by the user when completing this form. [3]

(Option C continues on the following page)

(Option C, question 9 continued)

Part of the file `process.php` is shown below.

```
<?php
$item = $_POST['item'];
$size = check_input($_POST['size']);

function check_input($data)
{ { if (!is_num($data){echo "Please check your selection";}
  else if($data>46 || $data<34) {echo "Sorry, nothing available";} }
return($data);
}
// other code present and not displayed

$link = $mysqli_connect(localhost, $username, $pwd, CatalogDB)
$sql = "SELECT it, sz, image, price FROM CatalogDB WHERE
       it = $item AND sz = $size";
$answers = mysqli_query($link, $sql);

if (mysqli_num_rows($answers) = 0)
  {echo "Nothing currently on stock matching your criteria.<br>"}
else
  {while($row = mysqli_fetch_assoc($answers))
    {echo "We offer " . $row["image"] . "<br>" .
      $row["it"] . $row["sz"] .
      " for " $row["price"] . "<br><br>"
    }
  }
mysqli_close($link)
?>
```

- (d) (i) Explain the processing that these two fragments of code enable on the server. [5]
- (ii) Outline an example of **one** script that could be added to the form and processed on the client's side. [2]

10. There are many discussion forums on “environment and climate change” on the internet. A scientific journalist has a blog, within the website of a major national newspaper, on this topic and other related topics.

- (a) With reference to the actions of “posting” and “commenting”, outline how a blog differs from a discussion forum. [4]

The articles in the blog are visibly tagged with keywords.

- (b) Discuss the use of tags in the blog in relation to web-indexing, by a web crawler. [5]

There are new discussion forums, on the same topics, created every day.

- (c) State **two** reasons why a new discussion forum may initially have a low search engine ranking. [2]

(Option C continues on the following page)

(Option C continued)

11. An assisted living project aims to provide elderly people with technologies that let them maintain some level of autonomy in their daily activities. This includes being supervised or monitored for healthcare purposes.

- (a) (i) Define the term *ubiquitous computing*. [1]
- (ii) Outline **two** examples in the above scenario where ubiquitous computing can be used. [4]

Information from these technologies is to be stored in the cloud in order to monitor the user's health from a distance.

- (b) Discuss the use of public and private clouds for storing this data. [5]
- (c) Outline **one** risk posed to elderly people from the use of these technologies. [2]
- (d) State the name of a protocol for mobile and wireless communication. [1]
- (e) Explain the importance of the use of protocols and standards in systems that support healthcare, in the context of this assisted living project. [5]
- 12.** (a) State **three** features that distinguish the semantic web from the web. [3]
- (b) Outline, with the use of an example, the difference between a web graph and a sub-graph. [2]
- (c) Define the term *multimedia-web*. [2]
- (d) Explain how a music search engine, which accepts audio input, functions. [4]

Peer-to-peer (P2P) networks are often used to download movies.

- (e) Discuss how power laws can be used to explain the phenomenon that movie files of the same size may take different lengths of time to download in a P2P network. [5]
- (f) Explain how collective intelligence affects the online buying habits of the general public. [4]

End of Option C

Option D — Object-oriented programming

A large company with locations in different cities has taken an OOP approach in creating an administration program that manages all aspects of its business. These aspects include:

- the sale of all of the different products that the company manages
- the salaries for managers, office staff and sales personnel.

13. (a) By making use of an example from the above scenario, distinguish between a class and an instantiation of a class. [3]

The different modules in the program each open up a graphical user interface (GUI). Each GUI has a similar design, but contains differences specific to each module.

- (b) By giving **two** examples, explain how the principles of inheritance can be incorporated into the design of this administration program. [4]

- (c) Describe how the use of libraries can facilitate the development of programs like this company's administration program. [3]

(Option D continues on the following page)

(Option D continued)

14. The company employs several sales personnel to sell its products to different retailers. Each branch of the company keeps track of its own sales with a suite of programs that include the two classes `SalesPerson` and `Sales`.

```

class SalesPerson // each object contains details of one salesperson
{
    private String id;
    private Sales[] salesHistory; // details of the different sales
    private int count = 0;        // number of sales made

    //constructor for a new salesperson
    public SalesPerson(String id)
    {
        // code missing
    }

    // constructor for a salesperson transferred (together with
    // their sales details) from another branch

    public SalesPerson(String id, Sales[] s, int c)
    {
        // code missing
    }

    public int getCount(){return count;}
    public String getId() {return id;}

    public void setSalesHistory(Sales s)
    {
        salesHistory[count] = s;
        count = count + 1;
    }

    public double calcTotalSales() // calculates total sales for the
                                   // salesperson
    {
        // code missing
    }

    public Sales largestSale() // calculates the sale with the largest
                               // value
    {
        // code missing
    }
}

```

Each instance variable is initialized when a `SalesPerson` object is instantiated.

- (a) Complete the constructor `public SalesPerson(String id)`, from the `SalesPerson` class. [2]
- (b) Explain why accessor methods are necessary for the `SalesPerson` class. [3]

(Option D continues on the following page)

(Option D, question 14 continued)

```
class Sales // each object contains details of one sale
{
    private String itemId; // id of the item
    private double value; // the price of one item
    private int quantity; // the number of the items sold

    // constructor missing

    public double getValue() {return value;}
    public int getQuantity() {return quantity;}
}
```

- (c) (i) Construct unified modelling language (UML) diagrams to clearly show the relationship between the `SalesPerson` and `Sales` classes. **Note:** There is no need to include mutator or accessor methods or a constructor. [4]
- (ii) Outline a negative effect that a future change in the design of the `Sales` object might have on this suite of programs. [2]

The company employs several sales personnel. The different `salesPerson` objects are held in the array `salesPeople`.

The `Main` class contains various methods that operate on the `SalesPerson` and `Sales` classes. The array `salesPeople` is declared globally. The `Main` class contains the following code:

```
SalesPerson[] salesPeople = new SalesPerson[6];
salesPeople[0] = new SalesPerson("100");
salesPeople[1] = new SalesPerson("101");
salesPeople[2] = new SalesPerson("102");
salesPeople[0].setSalesHistory(new Sales("A100", 300.00, 10));
salesPeople[0].setSalesHistory(new Sales("A200", 1000.00, 2));
salesPeople[1].setSalesHistory(new Sales("A300", 2550.40, 10));
System.out.println(salesPeople[2].getId());
System.out.println(salesPeople[0].getCount());
System.out.println(salesPeople[1].getSalesHistory(0).getValue());
System.out.println(salesPeople[0].calcTotalSales());
```

- (d) State the output after running this code. [4]
- (e) Construct the method `calcTotalSales()`, in the `SalesPerson` class that calculates the **total value** of the sales for a specific `SalesPerson` object. [5]

The `salesPeople` array contains 100 instantiated objects.

The company wishes to reward the salesperson whose sales have the largest total value.

- (f) By making use of any previously written methods, construct the method `highest()`, that returns the ID of the salesperson whose sales have the largest total value. [5]

(Option D continues on the following page)

(Option D, question 14 continued)

- (g) Construct the method `addSales(Sales s, String id)`, in the `Main` class, that will add a new `Sales` object `s`, to the salesperson with a specified ID.
Note: You can assume that the ID is a valid one. [4]

A further class in this suite of programs is the `Payroll` class. This class is run at the end of each month to calculate each salesperson's salary, which is based on the sales that have been made during that month.

- (h) Suggest changes that must be made to the `SalesPerson` class and/or the `Sales` class to allow these calculations to be made. [3]
- (i) Discuss the use of polymorphism that occurs in this suite of programs. [3]

15. The array `salesHistory` in the `SalesPerson` class is now replaced by objects of the `LinkedList` library class.

- (a) Identify the statement that will replace `private Sales[] salesHistory;` in the `SalesPerson` class. [2]
- (b) Construct the method `largestSale()` in the `SalesPerson` class, that returns the `Sales` object of the sale with the largest value. [5]

16. The following `Node` class defines a node to be used in a binary search tree.

```
public class Node
{
    Node left;
    Node right;
    int data;

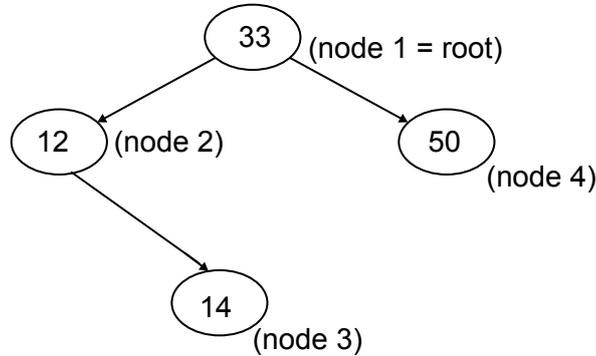
    Node(int newData)
    {
        left = null;
        right = null;
        data = newData;
    }
}
```

- (a) Using the above class as an example, describe the use of references as a data type. [3]

(Option D continues on the following page)

(Option D, question 16 continued)

The following binary search tree has been created using the data items: 33, 12, 14, 50. The nodes have been labelled for easy identification.



The method `print()` has been called with the statement:

```
print(root);
```

```
public void print(Node node)
{
    if (node == null) return;
    print(node.left);
    print(node.right);
    System.out.println(node.data + " ");
}
```

- (b) Using the data in the given binary search tree, trace the method `print()` showing clearly the recursive calls, the node that is being addressed and the order in which data is output. [5]

The method `smallest()` returns the smallest value of any binary search tree.

The method will be called by the statement: `int s = smallest(root);`.

- (c) By making use of the same `Node` class shown previously, construct the method `smallest()`. **Note:** You can assume that the data in each node is an integer and that the tree is **not** empty. [5]

End of Option D
