

Computer science
Higher level
Paper 2

Monday 7 November 2016 (afternoon)

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 – 9
Option C — Web science	10 – 13
Option D — Object-oriented programming	14 – 20

Option A — Databases

1. A college maintains a database to store information about the courses that it offers, as shown in the table below.

CourseTable

CourseID	Course_Name	Duration (Hours)	Cost
CS1200	Word Processing	10	30.00
CS1234	Databases	20	60.00
CS1201	Java Programming	105	150.00
CS5201	Java Programming - Advanced	50	90.00
CS1280	Python Programming	75	130.00
CS1331	C Programming	100	180.00
CS1209	Spreadsheets	20	30.00
...

- (a) Outline the difference between data and information. [2]
 - (b) Suggest an appropriate data type for each attribute in this database. [3]
 - (c) Explain the role of data validation. [2]
 - (d) State a validation check which might be helpful when entering a
 - (i) CourseID; [1]
 - (ii) Cost. [1]
 - (e) (i) Identify the steps in a query, that could be constructed to produce a list of the names of courses with a duration between 40 and 80 hours. [3]
 - (ii) Describe **one** function that is required to be performed on databases, other than the query function. [2]
2. The database schema is the definition that describes the entire configuration of a database.
- (a) Outline the characteristics of the **three** levels of the schema. [6]
 - (b) (i) Explain **two** responsibilities of a database administrator. [4]
 - (ii) Explain, with an example, how the database administrator would use a data definition language. [3]
 - (c) Identify **four** consequences of poor data modelling in the design of a database. [4]

(Option A continues on the following page)

(Option A continued)

- 3. Normalization can be described as a process of taking data from a given problem and designing it to a set of relations while ensuring *data integrity* and reducing *data redundancy*.
 - (a) Outline what is meant by data redundancy. [2]
 - (b) Outline what is meant by referential integrity of data. [2]

The following table, `College`, holds information about students, subjects and examination grades.

College

StudentID	Name	Gender	Subject	ExamGrade
CA3214	Ann Stern	F	Biology	5
CA3214	Ann Stern	F	Mathematics	5
CA3214	Ann Stern	F	Chemistry	3
SE1123	Mia Mann	F	Biology	7
FR2345	Joseph Hu	M	Computer Science	7
FR2345	Joseph Hu	M	Biology	6
FR2345	Joseph Hu	M	Physics	5
WR6543	Rea Kohn	F	Mathematics	4
...

- (c) Explain the steps that would have to be taken in order to hold this data in a relational database in 2nd Normal Form (2NF). You should include an example of the 2NF form in your answer. [6]
- (d) Describe **two** ways that a database management system (DBMS) can be used to promote data security. [4]

(Option A continues on the following page)

(Option A continued)

4. A hospital maintains computerized records. These records contain data about patients, doctors, medical facilities and treatment histories. The records are stored in an object-oriented database.

(a) Describe **two** characteristics of an object-oriented database. [4]

(b) Explain concurrency in data sharing in a database. [2]

This hospital is managed by a health care organization. The health care organization also manages 20 hospitals, over 200 clinics, a medical evacuation unit with several helicopters, and an insurance plan. They decided to create a data warehouse.

(c) Identify **four** operations that are used in the transformation process that integrates and normalizes data from multiple sources. [4]

The data warehouse is accessed by a variety of service providers within the organization including administrators, doctors, nurses, the medical evacuation unit and insurance companies.

(d) Describe **one** advantage of this warehouse for

(i) the patients; [2]

(ii) the service providers. [2]

(e) Data mining techniques could be used to uncover patterns regarding the health of the population.

(i) Compare classification and cluster analysis for this purpose. [4]

(ii) Suggest how deviation detection techniques could be used to increase the health care organization's knowledge about the population. [2]

End of Option A

Option B — Modelling and simulation

- 5. The population of rare tropical insects on an island follows a geometric model such that the population multiplies by k from one month to the next.

For example if $k = 2$ the population would double each month.

This situation can be expressed as either $P_{n+1} = kP_n$ or $P_n = P_0 k^n$.

P_0 is the initial population.
 k is a constant.
 P_n is the population after n months.

- (a) For a value of $k = 3$ and an initial population of 100, calculate the population after two months, P_2 . [2]
- (b) With the aid of a diagram, outline the way in which a spreadsheet could be used to estimate the size of the population after each of the first 10 months. [3]

Each month it is estimated that only 40% of insects survive due to various circumstances such as climate and lack of nutrition. This is known as a **constant survival rate** of 0.4.

- (c) Outline how this factor could be built into your spreadsheet, stating any assumptions that you make. [2]

Another model assumes a **changing survival rate** and recalculates the rate for each month using the following formula:

$x_{n+1} = rx_n(1 - x_n)$ where x_n is the survival rate in month n . ($0 < x_n < 1$)
 r is a constant.

For example, using $r = 3$ and a survival rate of 0.4 in the first month, the survival rate for the second month would be:

$x_2 = rx_1(1 - x_1) = 3 \times 0.4 \times 0.6 = 0.72$.

- (d) Construct an algorithm to calculate and output survival rates for the first 10 months, assuming that $r = 3$ and an initial survival rate of 0.4. [5]

The following table shows a comparison between a model with a **constant survival rate** of 0.4 and a model with a **changing survival rate**, for the first nine months.

Population (Constant survival rate)	100	64	51	41	33	26	21	17	13	11
Population (Changing survival rate)	100	72	61	71	61	71	61	71	62	71

- (e) With reference to the figures shown, suggest how these results could be used to compare the accuracy of the two models. [4]

(Option B continues on the following page)

Turn over

(Option B continued)

6. Influenza is a contagious disease and a new form is spreading throughout a city. Some people are affected more than others and the illness is spreading more quickly in some parts of the city than in others. Members of the public can be given a vaccine that will prevent them from contracting influenza. Unfortunately, a limited amount of vaccine is available.

(a) Outline how the characteristics of the people most affected could be determined. [2]

Age, gender and current health may be factors that affect risk.

(b) Describe how a model of those most affected could be defined and applied to the population to estimate the number of people who are at risk. [4]

(c) Outline **one** limitation of the model. [2]

Each day the areas in the city where people have contracted the illness are recorded. These records can be used to simulate an estimate of the areas immediately at risk.

(d) Describe the process of testing and refining the simulation. [3]

(e) Discuss how this modelling and simulation could be used to limit the spread and impact of the illness. [4]

7. Bridges are modelled and these models are tested using sophisticated software before being presented for construction.

The software enables the architect to:

- define the shape of the bridge
- apply parameters such as width, length and material to the structure
- view the bridge from different perspectives
- test the strength and stability of the bridge under all conditions.

The architect starts by defining the basic shape of the required bridge. This includes the span of the bridge (from one end to the other), and the struts (pillars) that are needed to support the bridge. During this phase the architect has a wire frame image to work with. When the design is complete, the image is rendered to give a solid view of the bridge.

(a) Outline the way in which the wire frame image can be used by the architect to adjust the parameters and view the bridge from many different perspectives. [3]

(b) Describe how the wire frame image of the bridge is held in memory. [4]

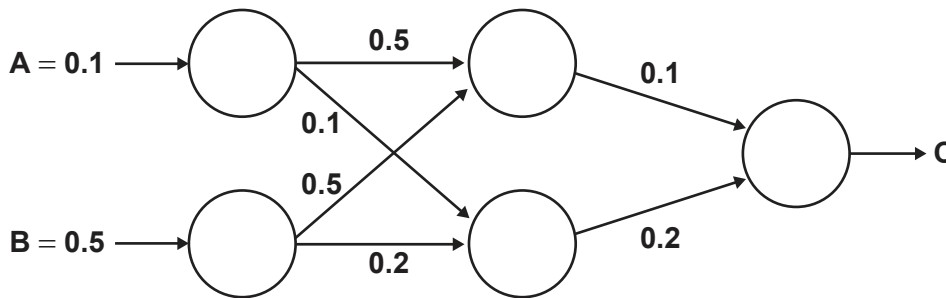
(c) Suggest reasons for making changes on the wire frame model rather than the solid view model. [3]

(d) Suggest a strategy to test the model and ensure that the bridge will be safe once constructed. [4]

(Option B continues on the following page)

(Option B continued)

8. The diagram below illustrates a simple feedforward neural network. A and B are the inputs and C is the output.



(a) With reference to the diagram, describe the use of the input layer and hidden layer in determining the output. You do **not** need to calculate the values. [4]

In optical character recognition, a genetic algorithm can be used to determine and refine the structure of a neural network.

(b) Explain how a genetic algorithm could be used to determine appropriate weights for this optical character recognition neural network. [5]

9. Research into natural language processing (NLP) has led to an improvement in machine translators.

(a) With reference to the text-to-text translation, outline the evolution of modern machine translators. [4]

(b) Describe the role of machine language learning in chatbots. [3]

With the increasing possibility of the use of robots in everyday life, research into human-robot communication is focusing on interpretation and repetition of movements and expressions so that robots can follow instructions.

(c) By considering the features of natural language, discuss how communication between humans and robots could benefit from interpretation of physical signs. [4]

End of Option B

Turn over

Blank page

Option C — Web science

10. A large file is being downloaded from a website using TCP/IP protocols.
- (a) Outline **one** difference between the internet and the World Wide Web (WWW). [2]
 - (b) Outline the function of the transmission control protocol (TCP) in the operation described above. [3]

The WWW is continually expanding and this has led to an upgrade in the internet protocol (IP). The IP that uses 32 bits is being replaced by one that uses 128 bits.

- (c) Outline the reasons for replacing the 32 bit protocol. [3]

The new IP uses a different format for representing the address field in the packet.

- (d) State the effect that this has on the domain name servers (DNS). [1]
- (e) Evaluate the choice of compression software in the transfer of the file. [5]

11. ACAME is a company that offers a subscription service that permits the downloading of animated cartoons. When browsing their website the following uniform resource locator (URL) is visited:

`http://minerva.hq.acame.net/Products/index.html`

- (a) Identify in the given URL
 - (i) the domain name of the server; [1]
 - (ii) the file path of the resource in the server. [1]

The web page contains the following line referring to a contact person.

`mailto:sly.coyote@acame.net.`

- (b) Outline why an expression related to mail services provides a uniform resource identifier (URI) but not a URL. [2]

Access to part of the website is restricted and requires authentication.

- (c) State **two** disadvantages of having the authentication based on the IP address of the client. [2]

ACAME has developed its website to be accessible on many different devices. CSS is used in the development of the website.

- (d) With reference to this use of CSS
 - (i) describe **one** advantage for the user; [2]
 - (ii) describe **one** advantage for the web developer. [2]

The website is extremely popular. ACAME relies on a large network of servers to store and distribute their cartoons.

- (e) Discuss the implications of having distributed systems. [5]

(Option C continues on the following page)

Turn over

(Option C continued)

12. A user interacts with a web form in order to select the towns where there are schools with computer science classes. The web form has two windows, one for the **country**, and one for the **town** where a school is located. The web form is made interactive to the user so that when selecting a country, only the appropriate towns in that country are displayed.

(a) Describe how this interaction can be generated dynamically. [3]

By making the selection through the web form, the individual websites of the selected schools, including other resources and services, may be accessed.

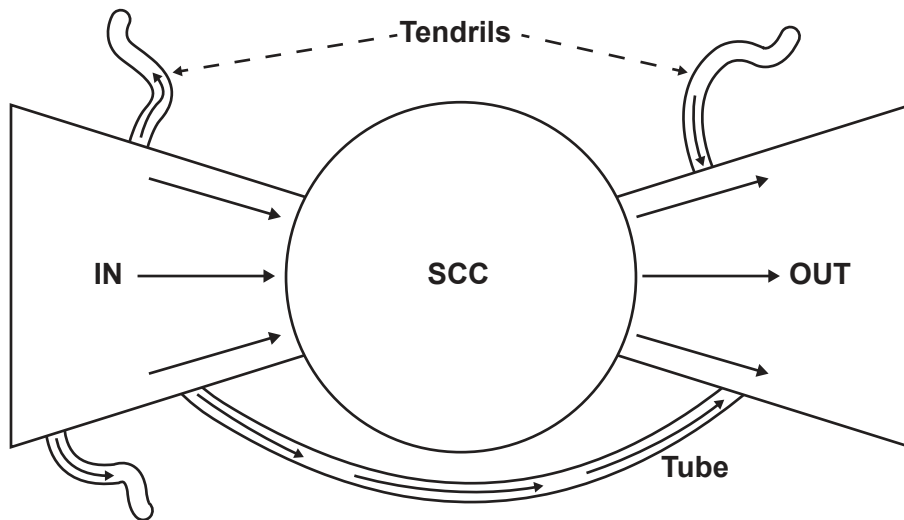
(b) Outline the role of open standards in providing web-based services. [4]

Schools manage their own website and resources, and the school website must be easily retrieved.

(c) Explain **one** example of the use of a white hat technique to optimize the search process. [3]

(d) Discuss, with an example, how a Creative Commons license could be used in this situation. [6]

13. The bowtie structure of the web is a representation of the navigational links among web pages, expressed with directed graphs.



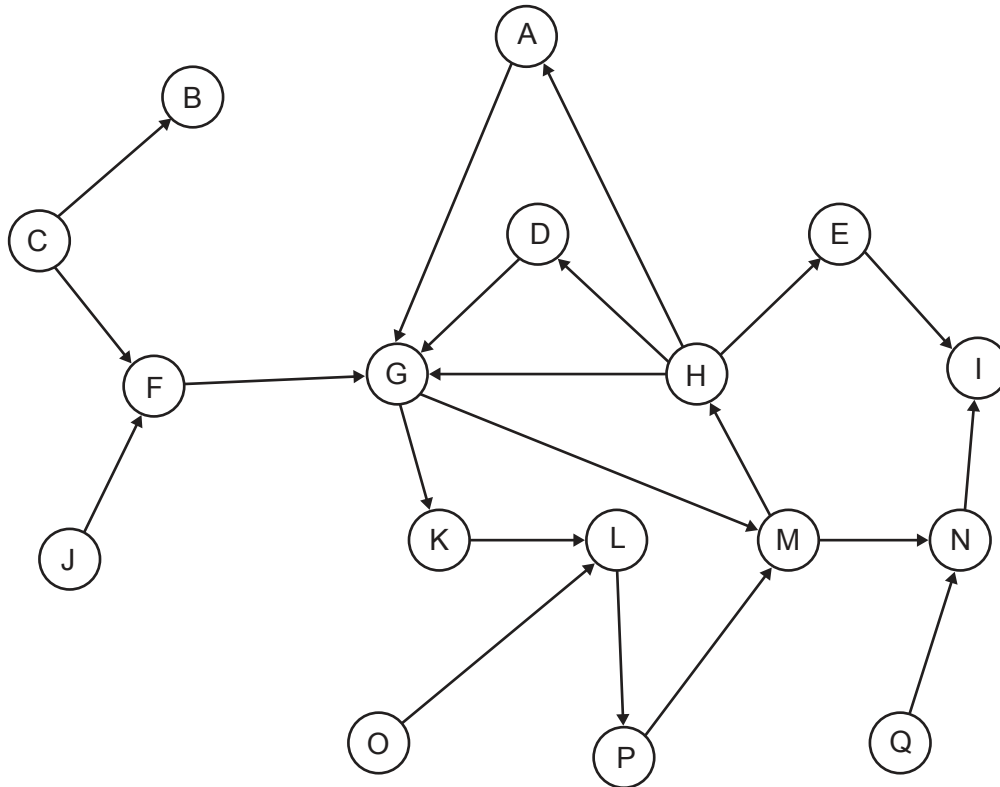
In the model some parts are identified, such as the strongly connected core (SCC), the IN and OUT of the web graph, the tendrils, and the tube.

[Source: adapted from A Broder, *et al.*, (2008), *Graph structure in the web*. *Computer networks*, **33** (1), pages 309–320]

(Option C continues on the following page)

(Option C, question 13 continued)

Consider the following graph.



- (a) Identify the nodes that make up the strongly connected core (SCC) in the graph. [2]
- (b) Identify the set of OUT nodes. [1]
- (c) State, giving a reason, whether the edge from J to F is a tendrill. [2]
- (d) Identify an edge that could be added to generate a tube, taking into consideration the existing nodes. Justify your answer. [2]

Search engines involve three main operations: crawling, indexing and ranking. By entering a keyword into the search field of a browser, websites are returned in some order of ranking.

- (e) Suggest **three** reasons that a web page may be inappropriately ranked following a search. [6]

Social networking has promoted the use of folksonomies.

- (f) Outline **one** disadvantage of the use of folksonomies in relation to the search process. [2]

The technologies and behaviours typical of the use of social networking have also contributed to the increase of scientific hoaxes and misinformation.

- (g) Discuss how collective intelligence can be exploited through social networking to promote an untested alternative therapy for a serious disease. [5]

End of Option C

Turn over

Option D — Object-oriented programming

In a particular country the Department of Transport analyses all of the train journeys of the different train companies in order to determine their reliability.

A team of programmers have created an OOP system for the Department of Transport to use. The `TrainCompany` class is partially shown below.

```
public class TrainCompany
{
    private String companyName;
    private String companyCode;
    private int numberOfJourneys;
    private Journey[] journeyHistory = new Journey[100000];

    public TrainCompany(String x, String y)
    {
        this.companyName = x;
        this.companyCode = y;
        this.numberOfJourneys = 0;
    }
    public TrainCompany(){}

    // accessor and mutator methods

    Journey getJourney(int x)
    {
        return journeyHistory[x];
    }

    public void addJourney(Journey j)
    {
        journeyHistory[numberOfJourneys] = j;
        numberOfJourneys++;
    }

    public double averageDelay()
    {code missing}
    // returns the average delay for all of a company's journeys

    public String longestDelay(Codes[] c)
    {code missing}
    // returns the route name for the journey with the longest delay

    public String toString(Codes[] c)
    {code missing}
}
```

(Option D continues on the following page)

(Option D continued)

14. (a) By making reference to OOP features, outline why it is possible for the `TrainCompany` class to have two constructors. [3]

(b) Outline the consequence of having the accessor and mutator methods declared as **protected**. [2]

15. Every time a train journey is completed a `Journey` object is created. The `Journey` class is partially shown below.

```
public class Journey
{
    private ... routeCode;
        // A unique identifier for a particular route
    private ... delay;
        // Minutes late in arriving at the destination
    private ... weatherRelated;
        // Equals true if the journey is affected by the weather,
        // otherwise false.
    // Constructor which initializes all 3 of the above data items
    // accessor and mutator methods.
}
```

(a) State the most appropriate data type for each of the three data items in the `Journey` class. [2]

(b) Construct the constructor method for the `Journey` class. [3]

The following code updates the journey information for each train company. This includes the creation of the data structure `allCompanies`, which is an array of `TrainCompany`.

```
TrainCompany[] allCompanies = new TrainCompany[3];

allCompanies[0] = new TrainCompany("Southern","T290");
allCompanies[1] = new TrainCompany("Northern","T400");
allCompanies[2] = new TrainCompany("Eastern","T155");

Journey s = new Journey("J100",3, false);
Journey t = new Journey("J103",8, true);
Journey u = new Journey("J104",10, true);

allCompanies[0].addJourney(s);
allCompanies[1].addJourney(t);
allCompanies[0].addJourney(u);
allCompanies[0].addJourney(new Journey("J101",6, false));

System.out.println(allCompanies[0].getCompanyCode());
System.out.println(allCompanies[0].getJourney(1).getDelay());
System.out.println(allCompanies[1].getNumberOfJourneys());
```

(c) State the output after the above code is executed. [3]

(Option D continues on the following page)

Turn over

(Option D continued)

16. The Department of Transport requires data from each company regarding the average length of delays. Delays that are caused by the weather are **not** included.

The `averageDelay()` method in the `TrainCompany` class returns the average delay for all of the journeys. The method does **not** count delays caused by weather.

- (a) State the value returned by the method call `allCompanies[0].averageDelay();`, assuming that the code in question 15 has been executed. [1]
- (b) Construct the method `averageDelay()`. You can assume that each `TrainCompany` object has at least one journey. [6]

17. A further class, `Codes`, links each `routeCode` to the `routeName`, which contains the departure and destination stations, as shown below.

```
public class Codes
{
    private String routeName; // e.g. New Amsterdam - Diamond City
    private String routeCode;

    public Codes(String a, String b)
    {
        this.routeName = a;
        this.routeCode = b;
    }
    public String getRouteCode()
    {
        return routeCode;
    }

    public String getRouteName()
    {
        return routeName;
    }
}
```

- (a) Construct the UML diagram for the `Codes` class. [3]

The array `allCodes[]` is an unordered array of `Codes` objects. It contains the details of every route for all of the train companies. The method `longestDelay()` in the `TrainCompany` class returns the route name for the journey with the longest delay not caused by the weather.

- (b) Construct the method `longestDelay()`. You can assume that all route codes are present in `allCodes[]`. [7]

(Option D continues on the following page)

(Option D continued)

18. The Department of Transport wants to expand this program to include similar analysis for bus and airplane journeys.
- (a) Describe how inheritance could be used to provide an organized set of classes for all of these forms of transport. [4]
 - (b) (i) Draw a diagram showing the dependencies between all of the classes described so far, including any introduced in part (a). You can assume that the `Journey` and `Codes` classes remain unchanged. [4]
(ii) Outline the reasons for keeping dependencies to a minimum. You should include an example using the classes described so far. [3]

The output from one call to the `toString()` method in the `TrainCompany` class is shown below.

```
Eastern : Average Delay = 2.5 minutes : Longest Delay = Newtown - Westlock Bay
```

- (c) Construct the `toString()` method. You should make use of previously defined methods. [4]
19. The programming team behind this system is considering making use of different data structures in order to improve efficiency.
- (a) Outline one reason why the array declarations in the original program may lead to inefficiency. [2]
 - (b) Currently the `allCodes` array is not structured for efficient searching. Describe how a **different** data structure can be both created and utilized so as to allow for the route name to be found more quickly. [5]
20. The programming team is considering whether to replace the `journeyHistory` array with an abstract data type (ADT) such as `ArrayList` or `LinkedList`.
- (a) Outline how a programmer is able to use methods of an ADT, such as an `ArrayList` or a `LinkedList`, without knowing how they are implemented. [3]
 - (b) Suggest why the team might choose to use an instance of the `LinkedList` library class, and not an `ArrayList`, to store the `journeyHistory` objects. [3]

The `allCodes` array has some duplicated entries (some objects which are the same). The team has decided to copy to an `ArrayList` object all **unique** entries of the `allCodes` array.

- (c) Construct the method `convert()` that will carry out this process. By inspecting the `routeCode` fields you should ensure that only unique objects are copied. [7]

End of Option D