Diploma Programme
Programme du diplôme
Programa del Diploma

Computer science
Higher level
Paper 2

Monday 6 November 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 – 9
Option C — Web science	10 – 14
Option D — Object-oriented programming	15 – 19

International Baccalaureate
Baccalauréat International
Bachillerato Internacional

Blank page

Option A — Databases

1. (a) Define the term *database transaction*. [2]

(b) Explain the importance of durability in a database transaction. [2]

(c) Identify **two** different types of relationships within databases. [2]

(d) Describe the nature of the data dictionary. [4]

(e) Identify **four** responsibilities of a database administrator. [4]

2. A group of art students have taken three examinations in the school year. Each of the three examinations has a maximum mark of 100. Each student's total mark for the year is the sum of the marks from the three examinations.

The students have passed unless their marks meet one or both of the following failing conditions:

- Scoring less than 30 marks on **any** one of the three examinations.
- Scoring a total mark that is less than 150.

Consider the following relation created by the teacher of this group of students.

CLASS_TABLE

Student_ID	Student_Name	Exam_One	Exam_Two	Exam_Three	Total
A1209	Ines Prest	30	39	30	99
A1422	Eva Maren	67	29	100	196
C4567	Joe Skrin	100	70	42	212
X0986	Perry Blare	54	68	30	152
B0078	Mia Lorres	26	100	100	226

(a) With reference to the relation CLASS_TABLE distinguish between data and information. [3]

(b) (i) Define the term *entity*. [1]

(ii) State the entity for this example relation. [1]

(c) Identify an appropriate data type for Student_ID. [1]

(d) Explain the role of data validation and data verification. [4]

(e) Explain how Total could be validated. [2]

(f) Describe the steps in a query that will output

(i) the names of all students who earned the maximum mark on Exam_Two. [2]

(ii) the Student_IDs of all candidates who passed. [4]

(Option A continues on the following page)

(Option A continued)

3. (a) (i) State what is meant by redundant data in databases.   [1]

(ii) Explain **one** issue that can be caused by redundant data in a database.   [2]

Consider the following example relation. It holds data about a number of teachers and students from different schools who volunteer to support the local community on particular days.

SCHOOL_VOLUNTEERS_TABLE (School_Name, Code, Address, Date, Num_Volunteers)

School_Name	Code	Address	Date	Num_Volunteers
Riverbank HS	AB1234	Goldmar Dr 15	10/07/2017	32
East High	EE1324	East Street 20	10/07/2017	15
East High	EE1324	East Street 20	20/10/2017	19
Oceanview HS	BG3445	Long Road 123	10/07/2017	14
Oceanview HS	BG3445	Long Road 123	20/10/2017	26
Blue Sky HS	SB9008	West Street 19	20/10/2017	37

The key attributes are underlined.

(b) Identify **three** characteristics of the 1st Normal Form (1NF) which are evident in this relation.   [3]

(c) Explain why a compound key is used for the SCHOOL_VOLUNTEERS_TABLE relation.   [2]

The following shows the normalized SCHOOL_VOLUNTEERS_TABLE relation:

SCHOOLS_TABLE
Code, School_Name, Address

VOLUNTEERS_TABLE
Code, Date, Num_Volunteers

(d) Discuss whether these relations are in third normal form (3NF).   [5]

(Option A continues on the following page)

(Option A continued)

4. A telecommunication company stores a large amount of data in three databases.

CUSTOMERS	A relational database which holds personal information such as age, gender, type of telephone, type of subscription, *etc*.
DETAILS_OF_CALLS	An object-oriented database which holds data about telephone calls such as call duration, call time, location, where call originated, call received, *etc*.
COMPONENTS	A spatial database which holds data about components of the telecommunication network, their location, their status and any error messages generated by these components, *etc*.

(a) Evaluate the use of an object-oriented database as opposed to a relational database. [4]

(b) Define a *spatial database*. [2]

The database segmentation is carried out on the CUSTOMERS database.

(c) (i) State what is meant by database segmentation. [1]

(ii) Explain **one** benefit of database segmentation to the telecommunication company. [2]

Data mining is used to extract knowledge hidden in this large amount of data. Before using data mining processes the existing data should be cleaned up.

(d) Explain how ETL processes could be used in data preparation. [3]

(e) Distinguish between the use of association and sequential patterns as data mining techniques. [3]

Customers committing fraud is a risk to the company.

(f) Describe how deviation detection could be used to detect fraud at the telecommunications company. [2]

Customers who decide to leave the telecommunication company for a competitor may result in huge losses for the telecommunications company.

(g) Explain with the use of an example, how predictive modelling could be used to optimize information sent to existing customers. [3]

End of Option A

Option B — Modelling and simulation

5. (a) Distinguish between a computer model and a computer simulation. [4]

(b) Identify **two** reasons why some systems are difficult to model successfully. [2]

Computer simulations are often used in situations where practical experimentation is, for some reason, not possible. One of these reasons could be an ethical issue.

(c) (i) With clear reference to the ethical issue, describe **one** example where practical experimentation would not be possible for ethical reasons. [3]

(ii) State **three** other advantages, apart from ethical reasons, of simulating a computer model rather than constructing a physical one. [3]

6. A manufacturing company that produces several products is using spreadsheet software to model its finances. This includes calculations that will estimate different quantities including the profit that the company will make in future years.

The model involves the use of spreadsheet software which will be organized using different sheets for different areas of the company's finances. Previously less sophisticated methods were used to keep track of costs and sales.

(a) By including examples where appropriate, describe a basic structure for this model. [6]

(b) Suggest how the reliability of the model could be tested. [2]

The company has established certain profit targets that it wishes to achieve over the next three years.

(c) Explain how this model can be used to investigate different strategies that will reach these targets. [4]

(Option B continues on the following page)

(Option B continued)

7. Weather forecasters use computer models which are able to simulate future weather patterns. These forecasts were originally limited to the near future. However, modern systems can now produce long range forecasts.

(a) Suggest **two** reasons why these simulations have improved both in their accuracy and their range. [4]

(b) Suggest why forecasts become less accurate the more long range they become. [4]

(c) Discuss whether historical data can be accurately used to forecast future weather. [5]

The simulation of the weather forecasting models produces specific data which can be output in a variety of ways.

(d) (i) Define the term *visualization*. [2]

(ii) With the help of examples, discuss how the development in the way such data is visualized has made the results of these simulations more accessible to the general public. [6]

(Option B continues on the following page)

(Option B continued)

8. Artificial neural networks (ANNs) are being used in a variety of applications, including pattern recognition, game playing, classification and data mining.

   (a) Explain **one** reason why ANNs are suitable for solving the types of problems found in these areas.   [3]

   An ANN is being trained to recognize handwritten numbers by identifying them as a digit from 0 to 9.

   (b) Describe an appropriate set of data that could be used to train the network.   [2]

   Assume that each digit to be identified is input as an image made up of 30×30 pixels.

   (c) State the number of neurons that would be in

      (i)   the input layer.   [1]

      (ii) the output layer.   [1]

   A simplified version of an ANN is shown below.
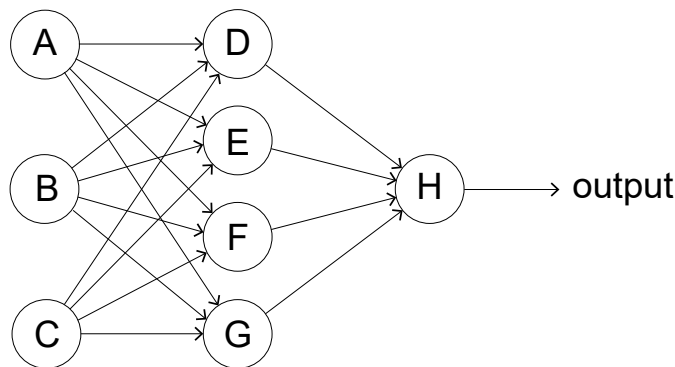
   (d) Explain the way in which the output from neuron D will be determined.   [3]

   The network is set up with initial values. The outputs are compared with the desired outputs.

   (e) Identify the steps that now take place to train the network.   [5]

9. *Google Translate* is an algorithm whose function is to translate text from one language to another. One of the resources that it uses is the body of documents produced by the United Nations which are routinely translated by humans into various languages.

   Discuss the reasons why *Google Translate* takes a probabilistic approach in preference to a cognitive rule-based approach.   [5]

End of Option B

Option C — Web science

10. *BuildYourWebSite* is an online company that provides a number of common templates for building your own website. Each template includes one HTML file, one CSS file, a folder of web images, and a folder of special sound effects.

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

(b) Discuss the benefits and disadvantages of the template including a CSS file in addition to the HTML page.   [6]

Each template can be downloaded as a single compressed file, using a web browser.

(c) Explain why TCP/IP is a reliable protocol in relation to downloading operations.   [3]

(d) Evaluate lossy compression and lossless compression when used to download files.   [4]

11. Open source code is made available by a community of developers and is frequently updated. The code can be downloaded for free, but users must register with the website and have their access authenticated.

(a) Evaluate the use of server-side scripting to provide the mechanism for registration.   [4]

Authentication is based on signing in to an established third-party company, for example a user's existing email or a social network account. The third-party company then verifies the user, granting them access to the open source code website.

(b) Explain how the user's privacy can be maintained whilst using this method of authentication.   [4]

The URL of this website is www.OpenSourceDev.org. Any new pieces of code that the developers make available become new resources on the website. A script generates weekly automatic notifications of new code available on the site, and sends this notification to users as an email.

(c) (i) Outline, with an example, how the URL for these new pieces of code will be generated.   [2]

(ii) Outline the steps that the script could perform for sending out these notifications.   [3]

(Option C continues on the following page)

(Option C, question 11 continued)

A fragment of a script and a web form are provided below.
In the script some functions are not implemented, and only their specification is provided.

```php
<?php
  //include a database of urls
  include('url_db.php');
  $url = $short = "";

  if ($_SERVER["REQUEST_METHOD"] == "POST")
  { $url = $_POST['url'];
    $short = make_short($url);
    function make_short($u)
    { $x = make_alpha_string($u);
      $y = first4_last4($x);
      $z = limits($x);
      $v = $y. ".".$z;  //string concatenation
      return $v;
    }

    function make_alpha_string($u)
    { // It removes, in this order: substrings corresponding to
      // protocol names, the substring www, and all characters
      // except for letters
    }

    function first4_last4($u)
    { // It returns the string made of the first 4 characters
      // followed by the last 4 characters of $u
    }

    function limits($u)
    { // It returns the string made of the first character and
      // last character of $u
    }

    mysql_query("
      INSERT INTO url_db(orig_url, short_url, url_ip) VALUES
      ( '".$_POST['url']."',
        '".$short."',
        '".$_SERVER['REMOTE_ADDR']."'
      )
    ");
  }
?>

<form method="post" action="">
URL:
<input  type="text" name="url" />
<br><br>
<input type="submit" name="Submit" value="Submit" />
</form>
```

(d)　Describe the processing that occurs when the form is filled with the URL
　　　https://www.the2nd.org/bin.php?id=70 that the server discovers is associated
　　　with the IP address 172.16.254.1.　　　　　　　　　　　　　　　　[3]

(Option C continues on the following page)

(Option C continued)

12. A museum has an online catalogue of its pieces. No new pieces have been acquired over the past 20 years, and the museum has already been at risk of closure for lack of funding. Its website consists only of static web pages.

It is suggested that the museum's static web pages may be contributing to the museum's lack of success.

    (a)   Identify **three** differences between a static web page and dynamic web page. [3]

    (b)   Suggest **two** services, and the benefits that they would provide, if the museum were to redesign its website to be dynamic. [4]

    (c)   Explain how the provision of services on the museum's website can increase its rankings in search engines. [4]

It is also suggested that the museum's basic catalogue be revised and upgraded with all of the museum's pieces being catalogued using meta-tags.

    (d)   Suggest how the use of meta-tags can be relevant for search engine optimization. [3]

**(Option C continues on the following page)**

(Option C continued)

13. With reference to the bowtie model of the web, consider the following graph. It represents a snapshot of popular sites about a specific topic.
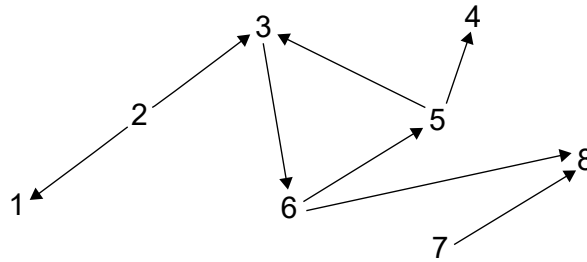
(a) Identify the nodes in the strongly connected core. [1]

(b) Identify the set of OUT nodes. [1]

(c) Outline why the edge between node 2 and node 1 is a tendril. [2]

After two months, the graph has evolved into the one shown below.
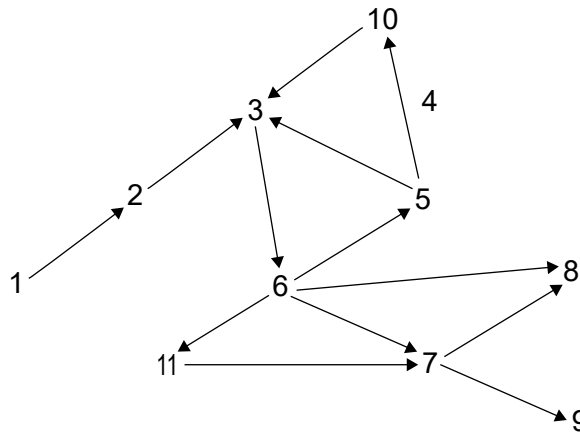
(d) Identify **two** reasons why an edge may disappear in the new graph, for example the edge between node 5 and node 4. [2]

We can interpret the graph as follows:
Each node represents a document and each outgoing arrow represents a hyperlink in the document.

(e) State how many hyperlinks are in the document corresponding to node 6. [1]

Power laws are often used to model phenomena occurring in very large networks and predict their evolution. Power laws may be suitably applied in different cases, in the context of the World Wide Web.

(f) With reference to the relation between the number of documents in the web and the number of hyperlinks in each of them, suggest how power laws may help to explain the shape of the World Wide Web. [5]

(Option C continues on the following page)

(Option C continued)

14. Smart technologies can be used to illuminate the night. Street light intensity and coverage can both be controlled, depending on physical parameters, such as the time of the day, or weather conditions.

    The amount of pedestrians in the street or their activities could be further parameters used by algorithms for these type of applications.

    (a) Discuss the technologies that ambient intelligence uses in this specific scenario. [4]

    (b) With a focus on privacy, discuss how traffic data from WiFi communication through smartphones can be used in this scenario. [4]

End of Option C

Option D — Object-oriented programming

A restaurant uses an object-oriented program to manage the cost of the food and drink consumed by customers. Everytime a table is occupied a `Payment` object is instantiated which will contain details of the items ordered. As each item is ordered, a `FoodItem` or a `DrinkItem` object is added to the `Payment` object as appropriate.

```
public class Payment
{
   private FoodItem[] fi = new FoodItem[100];
   private int fiCount;
   private static double foodTax = 0.2;   // 20% sales tax added to
                                          // all food prices
   private DrinkItem[] di = new DrinkItem[100];
   private int diCount;
   private static double drinkTax = 0.1; // 10% sales tax added to
                                          // all drink prices

   public Payment()
   {
      fiCount = 0;
      diCount = 0;
   }

   public DrinkItem getDi(int x)
   {
      return di[x];
   }

   // all other accessor and mutator methods are included

   // addFoodItem()  - this method adds a new FoodItem object
   // addDrinkItem() - this method adds a new DrinkItem object

   public static double findPrice(Item[] pl, String c)
   {  //code missing  }

   // calculateBill() - This method returns the bill (the total value of
   // the items consumed for a particular table)
}

public class FoodItem
{
   private String itemCode;
   private int quantity;

   public FoodItem(String x, int y)
   {
      itemCode = x;
      quantity = y;
   }

   // all accessor and mutator methods are included
}
```

The `DrinkItem` class is defined in a similar way.

(Option D continues on the following page)

(**Option D continued**)

15. Whenever a `Payment` object is instantiated, the variables `fiCount` and `diCount` are
     initialized to 0 through the code in the constructor.

     (a)   Outline an alternative method of initializing these variables that would not require the
           use of the code in the constructor.   [2]

     (b)   State the implication of the use of the term `static` in the `Payment` class.   [2]

     (c)   With reference to **two** examples from the classes on page 14, explain the benefits
           gained by the use of different data types.   [4]

     (d)   Describe the purpose of the following statement: [3]

           ```
           private FoodItem[] fi = new FoodItem[100];
           ```

     The `Payment` class method `addFoodItem()` is passed a `FoodItem` object as a parameter.

     (e)   Construct the method `addFoodItem()`.   [3]

(Option D continues on the following page)

(Option D continued)

16. The global variable `tables` is declared as follows:

```
Payment[] tables = new Payment[50];
```

The indices in this array represent the table number, so `tables[1]` is a `Payment` object for the customers occupying table number 1.

The driver (main) class contains the following code. **Note:** You can assume that all appropriate accessor and mutator methods have been included in their respective classes.

```
tables[1] = new Payment();
tables[2] = new Payment();
FoodItem a  = new FoodItem("f102", 2);
FoodItem b  = new FoodItem("f100", 1);
DrinkItem c = new DrinkItem("d102", 3);
tables[1].addFoodItem(a);
tables[1].addFoodItem(b);
tables[2].addDrinkItem(c);
tables[2].addDrinkItem(new DrinkItem("d103",1));
System.out.println(tables[1].getFiCount());
System.out.println(Payment.getFoodTax());
System.out.println(tables[2].getDi(1).getItemCode());
```

(a) State the output after this code is executed. [3]

(b) Construct statements, in code, that will print out the following:

  (i) The number of drink items ordered by table 40. [1]

  (ii) The item code of the third food item ordered by table 2. [1]

(Option D continues on the following page)

(Option D, question 16 continued)

The price of each item is stored in an object of the `Item` class.
The class is outlined below:

```
public class Item
{
    private String code;      // item code
    private String name;      // item name
    private double price;     // unit price before tax
    // all accessor, mutator and constructor methods are included
}
```

All of the objects in this class are held in the global array `pl` according to the following declaration: `Item[] pl = new Item[200];`
Note: The number of objects held in this array will change from week to week.

The method `findPrice(Item[] pl, String c)` in the `Payment` class looks up and returns the price of the item with code `c`.

(c)   Construct the method `findPrice()`. You may assume that the item exists in the array.   [6]

When a customer wishes to pay the bill, the `calculateBill()` method is called. If the bill was for table 10 then the following call would be made:

```
double finalBill = tables[10].calculateBill(Item[] pl);
```

(d)   Construct the `calculateBill()` method. You should make use of any previously
        defined methods.   [7]

17.   (a)   Construct a diagram showing the relationships between the `Payment`, `FoodItem`,
                `DrinkItem`, and `Item` classes. You do not need to include the names of attributes
                or methods.   [3]

        (b)   By making reference to any of the above classes, describe **two** benefits that a
                programming team would gain from the use of encapsulation.   [6]

The company that owns this restaurant also owns hotels, shops, and a car hire business.
The programs that manage the finances of these different businesses include classes similar
to the ones shown in this paper.

        (c)   Explain how inheritance could be used as a tool to both improve the clarity of the
                design and to reduce the amount of code that needs to be written.   [4]

(Option D continues on the following page)

(Option D continued)

18. The programming team have decided to replace the array of `Item` objects `(pl)` and the array of `Payment` objects `(tables)` with either a linked list or an ArrayList.

    (a) Outline **one** benefit that the use of either of these choices will have over the original array structures. [2]

The array of `Item` objects `(pl)` is replaced by `priceLL`, an object of the LinkedList library class. The `Item` objects can be created, deleted or amended.

The method `changePrice()` allows an item, identified by its item code, to have its price changed. All data required by this method are passed as parameters. If the item is not found then an appropriate message is displayed.

    (b) Construct the method `changePrice()`. [6]

    (c) Explain how a binary tree structure would allow a more efficient search for the `Item` object. [4]

19. A palindrome is a word that spells the same backwards. For example, the words "kayak" and "rotor" are palindromes.

A word can be checked to see if it is a palindrome by comparing the first letter with the last letter and then the second letter with the next-to-last letter and so on.

    (a) Explain why recursion is a suitable tool to use when performing this check. [3]

The method `palindrome()` is called by **boolean** `t = palindrome(word);` where `word` is a String variable.

    (b) Without writing code, describe the recursive method `palindrome()` that returns whether or not a word is a palindrome. [5]

End of Option D