Diploma Programme
Programme du diplôme
Programa del Diploma

Markscheme

May 2019

Computer science

Higher level

Paper 2

35 pages

International Baccalaureate
Baccalauréat International
Bachillerato Internacional

Subject details: **Computer science HL paper 2 markscheme**

Mark allocation

Candidates are required to answer **all** questions in **one** Option. Total 65 marks.

General

A markscheme often has more specific points worthy of a mark than the total allows. This is intentional. Do not award more than the maximum marks allowed for that part of a question.

When deciding upon alternative answers by candidates to those given in the markscheme, consider the following points:

- Each statement worth one point has a separate line and the end is signified by means of a semi-colon (;).

- An alternative answer or wording is indicated in the markscheme by a "/"; either wording can be accepted.

- Words in (…) in the markscheme are not necessary to gain the mark.

- If the candidate's answer has the same meaning or can be clearly interpreted as being the same as that in the markscheme then award the mark.

- Mark positively. Give candidates credit for what they have achieved and for what they have got correct, rather than penalizing them for what they have not achieved or what they have got wrong.

- Remember that many candidates are writing in a second language; be forgiving of minor linguistic slips. In this subject effective communication is more important than grammatical accuracy.

- Occasionally, a part of a question may require a calculation whose answer is required for subsequent parts. If an error is made in the first part then it should be penalized. However, if the incorrect answer is used correctly in subsequent parts then **follow through** marks should be awarded. Indicate this with "**FT**".

General guidance

Issue	Guidance
Answering more than the quantity of responses prescribed in the questions	• In the case of an "identify" question, read all answers and mark positively up to the maximum marks. Disregard incorrect answers. • In the case of a "describe" question, which asks for a certain number of facts *eg* "describe two kinds", mark the first two correct answers. This could include two descriptions, one description and one identification, or two identifications. • In the case of an "explain" question, which asks for a specified number of explanations *eg* "explain two reasons …", mark the first two correct answers. This could include two full explanations, one explanation, one partial explanation *etc*.

Option A — Databases

1.   (a)

> *Award [2 max].*
> *Award [1] for correct relationship "publishes books from".*
> *Award [1] for correct relationship "writes".*   **[2]**

(b)   *Award [2 max].*
As Book_Title attribute is String/Varchar;
It does not allow many validation checks (like range check *etc*);   **[2]**

(c)   *Award [1 max].*
Seeking the truth;   **[1]**

(d)   *Award [4 max].*
Award [1] for selection of Book_Title from Book
Award [1] for INNER JOIN Book ON PUBLISHER
Award [1] for JOIN Author.Author_Num = Book.Author_NUM
Award [1] for the correct test of equity of Publisher_Name.

SELECT Book_Title FROM Book
INNER JOIN Book ON PUBLISHER
Author.Author_Num = Book.Author_Num
WHERE Publisher_Name = "Orlando Crux"   **[4]**

(e)   *Award [2 max].*
Transaction must be completed in full;
Otherwise a rollback will occur;   **[2]**

(f)   *Award [2 max].*

Each user sees a fixed view of the data;
This may include visible changes made by the user / the user's own transactions
and transactions of other users

Changes made are not permanent and can be reversed using rollback;
This will return the database to its previous state (maintaining consistency);   **[2]**

(g)   *Award [2 max].*
When a field value changes multiple occurrences must be updated;
For example, if a publisher moves, we'll need to change the values for City and
Country in multiple records;
Problem occurs if we forget to change the values in any of the records;
The database would then have data inconsistency;   **[2]**

*Do not accept answer relating to storage requirements – it is not an issue in this
database*

2.   (a)   *Award [3 max].*
Row Locking;
Allows concurrent processing while ensuring transaction isolation;
Thus, ensuring the update of one teacher does not affect the update of another teacher;
While one transaction (by a teacher) is accessing a resource from a shared folder, it places a lock, an access restriction, on the resource, controlling the level of access allowed by another transaction by another teacher;   **[3]**

(b)   *Award [4 max].*
Award [1] for identifying a way of maintaining data security and [1] for a development up to [2 max]
Mark as [2] and [2]

Minimizing permissions/creating Access Rights;
Restrict users to have permission only to the means to do their job. Some can view, modify and insert some only view *etc*;

Auditing changes;
Log changes made to teachers and permissions through auditing. This gives a trail to follow should you have problems. Without authorization no one gets grant of permissions;

Minimizing table access;
Isolate the teachers from the data tables they do not need / Create views and user defined functions to support user access requirements and not give access to the tables;   **[4]**

Do not accept encryption as a response for this question

(c)   (i)   *Award [2 max].*
Award [1] for identifying a way of maintaining data security and [1] for a development up to [2 max].

Deferred update;
does not physically update the database on disk until a transaction has reached its commit point/if a transaction fails before reaching its commit point, it will not have changed the database in any way so UNDO is not required;

Shadow paging;
When a page is to be modified, a shadow page is allocated in which changes are made;
When it is ready to become durable, all pages that refer to original are updated to refer new replacement page;

Back-up / Mirror Copy;
Back-up/Mirror copies of the entire database are made to ensure the database is at the most updated version of the original;   **[2]**

Accept a log based recovery with description

(ii) *Award [3 max].*
Award [1] for the method, award [1] for the explanation and [1] for the example up to [3 max].

Data masking or data obfuscation;
is the process of hiding original data with random characters;
eg: suppressing certain characters in the student address, student id *etc*;

Data encryption;
Conversion of data into non-readable gibberish creates highly secure results such as scrambling the student_ID;
The only way to gain access to the data is to unlock it with a key or password which only those authorized can access; **[3]**

3. (a) *Award [4 max].*
Applications interacting with the database are minimally affected;
When a fully normalized database structure is extended, to accommodate new
types of data, database structure can remain largely or entirely unchanged;
So, the applications interacting with the database are minimally affected;

Key dependent;
Every non-key column in every table is directly dependent on the key, the whole
key and nothing but the key;
Remove redundant data;
Ensuring less anomalies and better efficiencies;
Mark as [2] and [2].   **[4]**

(b) *Award [2 max].*
For First Normal Form, each attribute value should be atomic;
In the given example product_Num, Pro_Name, Unit_Price and Qty is
multivalued / All 4 tuples shown have multiple values in their first 4 attributes;   **[2]**

(c) *Award [8 max].*
Award [1] for each correct table up to inc. appropriate foreign keys [4 max].
Award [1] for each correct primary key up to [4 max].

3NF

Example 1

Product
(<u>Product_Number</u>, Unit_Price, Product_Name)

Salesperson
(<u>Sales_Person_</u>Number, Sales_Person_Name, Manager_Number)

Manager
(<u>Manager_Number</u>, Manager_Name)

Purchases
(<u>Purchase_ID</u>, Date_And_Time, Product_Number, Sales_Person_Number)

Example 2

Product
(<u>Product_Number</u>, Unit_Price, Product_Name)

Salesperson
(<u>Sales_Person_Number</u>, Sales_Person_Name, Manager_Number)

Manager
(<u>Manager_Number</u>, Manager_Name)

Purchase
(<u>Date_And_Time</u>, <u>Product_Number</u>, <u>Salesperson_Number</u>)   **[8]**

(d)   *Award [2 max].*
        You can enter foreign keys that do not match the corresponding primary key in
        the related table;
        This could cause a lot of problems such as mismatched customer data and
        mismatched transaction records;

        Cascading update;
        If the primary key for a record in the Managers table changes, all corresponding
        records in the Employees table are modified;

        Cascading delete;
        If a record in the Managers table is deleted, all corresponding records in the
        Employees table are deleted;   **[2]**

(e)   *Award [2 max].*
        The combination of the two provides unique identifiers for the records of the
        table;
        And that there is no single attribute that is able to uniquely identify a record;
        In a manner that will not lead to potential duplication of records;
        It identifies exactly one record of the table, then that record shows the single
        value of each of the non-key attributes;
        That is associated with the unique combination of the key attributes;   **[2]**

4. (a) (i) *Award [2 max].*
Data warehouses contain both historical and current data;
Timestamps are required to compare data from different times; **[2]**

(ii) *Award [2 max].*
Increased Query and System Performance;
data warehouse is built for analysis and retrieval of data rather than
efficient upkeep of individual records (*ie* transactions);

Timely Access to Data;
ETL, are used within a data warehouse environment. These routines
consolidate data from multiple source systems and transform the data into
a useful format that enable quick querying;

Enhanced Data Quality and Consistency;
Data from the various business units and departments is standardized and
the inconsistent nature of data from the different sources is removed;
Individual business units will start to utilize the same data repository as the
source system for their individual queries and reports;

Historical Intelligence;
Data warehouse stores large amounts of historical data and time-period
analysis, trend analysis, and trend prediction thus allowing for advanced
reporting and analysis of multiple time-periods; **[2]**

(b) *Award [2 max].*
When the data is collected from different sources each source will have their own
standards if we have two different data sources A and B;
Selection of the data that is going to be useful in analysis – the different offices
will have data relevant only to them, such as staff names, *etc*;
Standardization of the data – the company may have imposed a standard on all
its offices, but this is not always the case, and the data will certainly have to be
Checked, *eg* date formats are different in different countries;
Other transformation techniques that students may give as examples are more
like:

- Cleaning (*eg* "Male" to "M" and "Female" to "F" *etc*)
- Filtering (*eg* selecting only certain columns to load)
- Enriching (*eg* Full name to First Name , Middle Name , Last Name)
- Splitting a column into multiple columns and vice versa. **[2]**

(c) *Award [6 max].*
Award [1] for a description of cluster analysis
Award [1] for description of cluster analysis being used to find patterns
Award [1] for a description of link analysis
Award [1] for a description of link analysis being used to find patterns
Award up to [2] for comparison between the two techniques

Cluster Analysis
Cluster analysis groups customers by age / according to different factors, such as region/location;
Therefore, it enables comparison between groups;

Classification
A classifier or model is developed using training sets of data;
New data is then added to the model and compared against the predicted outcomes / new classifiers may be developed;

Discussion
Classification requires prior knowledge of the customer base, cluster analysis does not;
Data can be classed in new samples using classification whereas for cluster analysis ~~one~~ only suggests groups based upon patterns in data;
Labelled samples from a set of classes is required for classification whereas for cluster analysis unlabelled samples will do;   **[6]**

(d) *Award [3 max].*
Access the customer payment details for purchases of fairly large orders from the data warehouse;
For a given group (and for a particular period/given a range for timestamp) of customers;
Identify the outliers/customers who have defaulted more than a fixed number of times in payment after running the deviation detection (multiple if-then-else statements) algorithm;
List the names of those customers;   **[3]**

(e) *Award [5 max].*
Award [1] for an outline of customer segmentation
Award [1] for an outline of link analysis
Award up to [3 max] for use by ZCC to improve marketing strategies

Customer segmentation
Divides a customer base into groups of individuals that are similar in specific ways relevant to marketing, such as age, gender, order of the type of paper, frequency of orders placed and the size of the orders normally placed;
assign each customer to one of the segments;
A typical segmentation makes each segment distinct from other segments (different segments have different needs), it is homogeneous within the segment (exhibits common needs);
segmenting is using borders to form groups;
Segmentation groups objects into similar groups;
The resulting groups contain members that are more similar to each other than they are to other groups;

Link Analysis
Load a claim and runs a query back to the database to find all other claims sharing any similar attributes;
show matches on the address of a claimant in the original case being investigated;
Now combine matches-merge identical nodes. So, we can more easily see unusual connections;
Once seen a suspicious link, accept or escalate;
Representing data as a network offers an engaging way for analysts to rapidly understand events; **[5]**

Option B – Modelling and simulation

5. (a) *Award [3 max].*
Award [1] for each correct row

Variable	Data type	Range of values
Ocean surface temperature	Number, real	Between –5 and +30
Albedo	Number, real	0–100 or 0.1 - 0.9
Sea ice area	number	Accept any reasonable range such as 0 - 2 000 000

[3]

(b) *Award [4 max].*
Initialization **–** Original area of sea ice;
Initialization – Surface temperature change;
New sea ice area calculation;
Sea level change calculation;

Example algorithm:
```
OriginalSeaIceArea = 1000000
SurfaceTempChange = 0.04
NewSeaIceArea = OriginalSeaIceArea * (1 – SurfaceTempChange)
SeaLevelChange = (NewSeaIceArea - OriginalSeaIceArea) /
OriginalSeaIceArea * 100 * 20
```

Do not penalise for using other methods to calculate correct values. **[4]**

(c) (i) *Award [1 max].*
960 000 (km^2); **[1]**

(ii) *Award [1 max].*
80 (mm); **[1]**

(d) *Award [7 max].*
Initialization – Initial areas of Ice and water, and starting year;
Use of loop;
Loop parameters (limits, condition, increment, end statement correct);
Average albedo;
Rate of decrease calculation;
Area of ice update calculation;
Area of open water update calculation;
Updating/increasing of year by 2;
Correct output;

Example algorithm:
```
IceArea = 1000000
OpenArea = 1000000
Year = 2019
Loop While IceArea >= 10000
   TotalArea = IceArea + OpenArea

AveAlbedo = ((IceArea * 0.6) + (OpenArea * 0.1)) / (IceArea + OpenArea)
   Decrease = 0.3 / (AveAlbedo * AveAlbedo)
   // allow Decrease = 0.3 / AveAlbedo ^ 2
   IceArea = IceArea * (1 – (0.01 * Decrease))
   OpenArea = TotalArea - IceArea
   Year = Year + 2
End loop
Output "Date when there less than 10 000km² of ice in the ocean
is " Year
```
[7]

(e) *Award [2 max].*
Spreadsheets;
IDEs using code;   **[2]**

(f) *Award [3 max].*

Example Answer 1:
If the calculation of the rate of reduction of the sea ice is inaccurate the
calculation of the amount of sea ice remaining;
At a given time will turn out to be wrong/incorrect;
So, work to correct the problems may be too little/too great;

Example Answer 2:
The year in which the sea ice is predicted to be 10 000 km²;
Will be wrong/incorrect;
So, ice coverage may be more/less than expected in the calculated year;   **[3]**

6. (a) *Award [4 max].*
Data is entered into a rule-based environment;
The rules may be kept the same and the nature of the data input may be varied;
Or the data that is input may remain the same and the rules may be varied;
Which allows for a range of possible scenarios to be investigated; **[4]**

(b) *Award [3 max].*
Life expectancy;
Gender;
Availability of life-extending treatments;
Genetic information;

Allow other appropriate answers (such as profession/family health history/income/social-economic condition/inclination to sporting activities). **[3]**

(c) *Award [5 max].*
Award [1] for identification of an issue and up to [3] for elaboration; Candidates must come up with 2 issues or of similar nature but leading to different consequences.

One ethical issue is invasion of privacy;
to get the best quality model;
sensitive personal data will need to be collected;

trade-off between the benefits of this model and level of intrusion into people's lives;
especially if the data is being shared with third parties; **[5]**

(d) *Award [3 max].*

Example Answer 1:
Simulation may be used to handle uncertainty and provide ranges of expected outputs;
e.g. repeatedly inputting data drawn from random samples of plausible input values;
To look at the predicted spread of outcomes over time (consider the use of simulation in the analysis of queues);

Example Answer 2:
A simulation is considered to be a representation of a model over time;
This means that the simulation can be used over a longer period of the development life cycle;
It can be used to assist with stages such as the implementation stage; **[3]**

7.   (a)   The creation of a human interpretable image/graph from data;
              Allow any sensible definition that defines the term.   **[1]**

       (b)   *Award [2 max].*
              The characters are defined as mathematical models;
              They will need rendering to give a 3D effect;
              Each movement will need to be fluid;
              The animation is less interesting with flat figures;

              Allow [1 max] for only a definition of rendering.   **[2]**

       (c)   *Award [6 max].*
              *Award [1] for identification of implication and up to [3 max] for the development of
              the implication identified.*

              Rendering for non-interactive media such as animations is a much <u>slower
              process</u> than for interactive media/games;
              Rendering times for individual frames may vary from a few seconds to days
              depending on the complexity of the scene;
              Rendered frames are stored to a hard disk and then transferred to other media
              for playback;

              3D rendering is <u>computationally resource hungry</u>;
              Computer processing power has increased rapidly and greatly over the years;
              Allowing a much higher degree of photorealism / realistic rendering;
              However, the enough computational power still can't cope up with growing
              demand of realistic 3D animation.   **[6]**

8. (a) *Award [4 max].*
A genetic algorithm works in the same way as an evolutionary process whereby it starts with a large population;
And uses an iterative process where the fitter solutions are selected and input into the next cycle until the exit criteria are satisfied;

Whereas neural networks attempt to mimic the process of the brain;
And can be used/trained to recognize patterns;   **[4]**

(b) *Award [2 max].*
Increase the number of inputs;
Increase the number of hidden layers;   **[2]**

(c) *Award [4 max].*
Supervised learning is when the outcome related to a given input is already known;
And so, the "learner" can recognize objects and name them based on the labels already given;

Whereas unsupervised learning is when no examples of outcome are given to help with the learning;
And so, the "learner" must deduce its own solutions *eg* classifying similar objects by colour or shape;   **[4]**

(d) *Award [4 max].*
The language of the child may not have been programmed into the doll;
The child may not speak clearly;
The child's language may not be sufficiently developed to apply syntax correctly;
The child may refer to something not in Alicia's "recorded" content, or that Alicia has not previously "learnt";   **[4]**

(e) *Award [6 max].*
Unsupervised learning can be used for bridging the causal gap between input and output observations;
Instead of finding the causal pathway from inputs to outputs;
By building the model upwards from both sets of observations;
In the hope that the gap is easier to bridge in the higher levels of abstraction;
Possible to learn larger and more complex models;
e.g. the connection between two sets of observations;   **[6]**

Option C – Web science

9. (a) *Award [2 max].*
Change/re-flow the layout of the page to suit different screen sizes/formats;
adjust font sizes;
adjust image sizes;
provide alternative menus / link options;
It can detect the device and adjust page size accordingly;

(b) *Award [3 max].*
The code:
- uses SQL to query the CATEGORY table in a database, selecting all pairs of `<category_id, category_name>` (ordered by `category_name`);
- stores the results of the query in a variable / array `$categoryqueryresult`;
- prepares HTML for a drop-down menu which consists of a list of option tags derived/obtained from iterating/looping/going through each pair in `$categoryqueryresult`;
- sets the value of each item to `category_id` and text to `category_name`;
- creates a form whereby the result is sent to "`showresults.php`" using the POST method

(c) *Award [4 max].*
Mark as [2] and [2].

Include authoritative / high quality / unique website content;
To attract other reputable sites to link to it / drive traffic to the site;

Use appropriate/ descriptive meta tags (*e.g.* keywords/descriptions);
To provide clear data for web-crawlers/robots to use when categorising the page;

Separate content from formatting (e.g. use of CSS etc.);
To allow search engines to index the content of the site more effectively;

Include a robots.txt file in the page header;
To give instructions to web-crawlers/robots as to how to index and describe the various pages on the site;

Link baiting, creating content for your site / other popular sites that aligns with your site but has a different angle (e.g. guest blogging);
that will compel others to create links to / visit your site;

(d) *Award **[2 max]**.*
*Mark as **[1]** and **[1]***

Lossy compression
Reduces the file-size by removing some of the data in the file;
Once removed, the data cannot be recovered;
This generally results in a loss of quality (*e.g.* picture resolution, audio frequency range);
Files compressed using lossy compression are used in their compressed form (*e.g.* images, video, audio);

Lossless compression
Reduces file size by looking for repeated patterns of data / redundant data and replacing those with a single shorter "token"; *//do not penalize the candidate for not using the word "token"*
Files must be decompressed before they can be used/ to recover the data;
Files do not lose any of the data/quality they contain when compressed / decompressed;

(e) *Award **[3 max]**.*
Lossless compression reduces file size while preserving all the data;
Once decompressed, the pdf document will contain all the data in its original form;
It would be hard for a lossy compression algorithm to distinguish between data that can be removed (*e.g.* reducing the resolution of the images) and data that is necessary for the pdf documents to make sense (*e.g.* parts of the text);
Using lossy compression may result in an unusable/unreadable document;

Accept answers that focus on why lossy compression would be unsuitable for the pdf files.

10.   (a)   (i)   *Award **[1 max]**.*
                     HTTPS / hypertext transport protocol secure;

          (ii)   *Award **[3 max]**.*
                     The DNS looks up the domain name "home.cern" in its database;
                     If it doesn't have this string, it passes the query to another DNS according to defined rules;
                     This process continues until either an IP address is passed back to the starting DNS or an error message is returned;
                     The IP address (or error message) is sent back to the client that initiated the call to the DNS;

      (b)   *Award **[6 max]**.*
            *Mark as **[3]** and **[3]***

            Multiple copies of all or part of the data can be kept at different sites;
            This ensures that there is no single point of failure/ increases reliability;
            and the redundant data helps to ensure against data loss;

            Different computers on the grid can use different analysis and data visualization tools;
            This allows scientists to run whatever analysis tools best suit their own specialism/area of interest;
            Rather than being limited to the tools provided by CERN;

            Analysis can be performed using distributed processing time/capacity;
            This reduces load and/or reliance on a centralized system / speeds up processing;
            Allowing a greater number of processes to be run concurrently;

            Computers on the grid can be in multiple time zones;
            This gives scientists more equitable access to data;
            And facilitates round-the-clock monitoring and the availability of expert support;

            Resources can be distributed across the world rather than being held in one country;
            This may attract funding from governments for their own locally-based research;
            As they may see the benefits of international cooperation;

            Grid computing is scalable / extensible;
            Simply add/remove more resources to increase/decrease computing power as required;
            With little complexity / cost;

(c)   *Award [6 max].*
        Mark as [3] and [3]

Reason:
CERN want their experimental results to be freely accessible (within specified limits);
Development of reason:
Allows for the more rapid dissemination of data/information
May further the advance of scientific knowledge / be seen as an altruistic gesture;
No need to contact CERN about using the work / allows CERN to focus on their primary function, *i.e.* scientific research;
prevents people from repackaging them and selling them as a commercial product;
Results may be used without being worried about being plagiarised;

Reason:
Enforcement of copyright would require significant costs;
Development of reason:
Associated with hiring of lawyers;
May be impossible to enforce;
It may not be possible to find all cases where work has been used without copyright permissions;
Infringement of copyright may occur outside of Switzerland where different copyright laws may exist;

Accept other reasonable answers

11. (a) *Award [1 max].*
         The ability of something to be extended or expanded from its initial state, *eg*
         software / file Formats / programming languages;
         The ability to add custom / user-defined elements (*eg* XML tags, plug-ins,
         add-ons);

      (b) *Award [2 max].*
         In the real world, computer systems and databases contain data in incompatible
         formats / structures;
         XML data is stored in plain text format. This provides a software- and hardware-
         independent way of storing data / makes it much easier to create data that can be
         shared by different applications;
         XML makes it easier to expand or upgrade to new platform / operating systems /
         applications / browsers;
         Separates the data from HTML;

      (c) *Award [2 max].*
         Interoperability is the ability of different computer systems (including operating
         systems and applications) to work cooperatively / share data / exchange data;
         Open standards are standards that are publicly available and (normally) free to
         use;
         Open standards are one factor aiding interoperability;

      (d) *Award [2 max].*
         The Common Gateway Interface (GGI) is part of the Web's Hypertext Transfer
         Protocol (HTTP);
         CGI is a method or convention for passing data back and forth between the
         server and the application;
         CGI provides a more efficient mechanism for data to be passed from the user's
         request to the application program (and back to the user);
         CGI is not dependent on the operating system used by the server. The methods /
         conventions remain the same;

      (e) *Award [2 max].*
         In a blog, only the owner can post an article / open a thread of discussion / start a
         theme;
         In a forum all registered participants can post an article / open a thread / start a
         discussion;

         In a blog, registered visitors may be allowed to comment but the blog owner may
         moderate the comments before displaying them;
         in a forum all comments from registered users are published immediately
         (moderation may happen later);

(f) *Award [6 max]*.
   Note: do not award a tick for each idea, use the markband and use best-fit to determine the level of the response.

Blog
- allows the head librarian to post articles/entries that give readers a better idea of the issues/decisions affecting the library/the background context that affect decisions / may result in better informed and focused comments from readers;
- may foster a greater sense of participation / spirit of community / allow the head librarian to aggregate majority views;
- the ability of the head librarian to moderate posts may allow off-topic/ unhelpful/offensive comments to be filtered-out / however this might also allow the head librarian to censor/modify comments that are legitimate but critical of library services/contrary to decisions that the library want to push-forward;

Forum
- allow all library users to raise issues that they find important rather than only commenting on issues raised by the head librarian / may allow head librarian to develop a greater awareness of issues affecting/affected by decisions;
- lack of moderation may allow users to raise controversial issues / however may also reduce ability of head librarian to filter-out off-topic/unhelpful/ offensive posts;
- may oblige head librarian to engage in greater discussion / justification of actions / however excessive time might be spent clarifying issues / dispelling myths;

Both blog and forum
- the comments/posts may not be reflective of the general/majority view / may be restricted to a biased/self-selecting sub-set of library users. This may influence the head librarian to make decisions that cater for the "vocal" minority;

[1–2 marks]
A limited response that indicates very little understanding of the topic or the reason is not clear. No reference is made to the scenario in the stimulus material. The response is theoretical and descriptive.

[3–4 marks]
A superficial analysis of why the increased engagement with library users through either the blog or the forum or both will assist the head librarian with decision making. Some references are made to the scenario in the stimulus material.

[5–6 marks]
A discussion of why the increased engagement with library users through the blog and forum will assist the head librarian with decision making. Explicit and relevant references are made to the scenario in the stimulus material.

12. (a) (i) *Award [1 max].*
D, E, F, G;

(ii) *Award [1 max].*
H & K;
Accept H & L & K;

(b) *Award [2 max].*
Both pages have the same number of in-links (*ie* 2);
However, the links to page E come from pages that have a greater number of in-links than the pages that link to C / E is connected with SCC while C has two in-nodes / in-links;
The PageRank algorithm counts links to pages recursively;
A PageRank algorithm will give a greater weighting to the pages that link to E;
Therefore, the algorithm will place E higher up the ranking than C;

(c) *Award [3 max].*
Whether there are any meta-tags present that restrict / guide indexing (*eg* a "robot exclusion protocol");
Whether there is a robots.txt file linked to the page that gives instructions to the web crawler;
Whether the page has broken / dead links;
Whether the page content / meta information matches any specialism/type sought by the web crawler (*e.g.* some crawlers specifically target academic content);
There is no header with meta-tags;
Whether the page has ever been indexed before;
Whether the page has changed since it was last indexed;
How long ago / how frequently the page has been indexed (web-crawlers will tend not to index pages too frequently as this increases load on the web server);

(d) *Award [5 max]*

Reasons why power laws may be appropriate
The number of users shown for each site appears to follow the general pattern of a power law distribution / is consistent with the general principles of power laws (the "rich get richer");
It's likely/reasonable to assume that sites with large numbers of users will tend to attract more new users than sites with fewer users;
This may be particularly true of social media sites where a high number of existing users may equate to a more diverse/engaging/attractive experience for new users;

Reasons why power laws may not be appropriate
However, correlation does not equal causation. / Just because the sites appear to exhibit a power law distribution, it doesn't mean that their growth is governed by a power law;
Other factors may be more significant (*e.g.* the demographic / region a social media site attracts, changing fashion, the policies of the sites themselves);
Sites that were very popular in the past but diminished/died-out may suggest that power laws are not the only / main factor influencing future development (*e.g.* MySpace);

13.   (a)   *Award [2 max].*
Enables Narayan City to gather data from many more sources than would be possible through the use of traditional data-gathering methods;
This will enable far more information to be gathered far more quickly than using traditional methods;

Data is generated proactively by citizens' actions;
Therefore, data more likely to reflect what citizens actually do rather than what they might claim they do;
Improves the accuracy of the data;

Existing strategies used by citizens can be incorporated into future planning;
for example, using side-roads to optimize journey times/timing of journeys to avoid major congestion;
Thus, helping to ensure that future plans are optimized / tailored to citizens' needs;

Data gathering and analysis can be automated;
Therefore, it is very cost effective;
Thus, it is able to be scaled / increased without major impacts on budget;

Data monitoring and collection is real-time;
Therefore, the effect of implementing plans will be reflected by visible changes in use patterns / journey times / routes taken;
Thus, the city council can get reliable feedback/evaluation of how effective their plans are;

(b)   *Award [6 max]*.

Possible changes in user behaviour
- users can create their own hashtags, allowing the opportunity for issues/topics to be classified that subsequently rise in prominence/trend among a wider range of social media users/increases the chance of users' concerns being noticed and addressed;
- contrasting/conflicting hashtags can spark debate and discussion among users, allowing users to be exposed to perspectives and arguments that they would otherwise miss / be unaware of;
- Users interact with content related to their interests / views by following hashtags / social media accounts they know align with their interests / views
- Hashtags / social media accounts provide opportunity / platform / communication channels to connect with people for collaboration
- Folksonomies / hashtags / social media / web 2.0 has given all users opportunity to create instead of simply seeking information;
- trending hashtags may make users with opposing points of view less willing to expose their dissent / more likely that users just follow the trend;
- Social media allows people to be anonymous and hence more likely to attack/hurt others
- users including certain hashtags may find themselves open to attack/abuse from other human users / automated bot accounts (*e.g.* "trolling");
- hashtags are open to misuse and manipulation (*e.g.* "bot" accounts flooding social media with tagged posts / users stuffing posts with multiple lines of the same hashtag to create a false "trend");
- users who follow/search on popular hashtags may find themselves "bubbled" – *i.e.* they are only ever exposed to viewpoints that reflect their own existing perspectives;
- Content people share is moving towards viral / inappropriate content to attract attention
- Followers of popular accounts / hashtags tend to be influenced by the views / content of the account / hashtag (e.g. popular YouTube accounts may start advertising products)

Accept other reasonable answers that convey similar ideas.

Option D – Object-oriented programming

14. (a) *Award [1 max].*
A method that makes/allows/controls changes to a private (hidden) / instance variable / variables in an object or class;   **[1]**

Note to examiners: as this is OOP, some reference must be made to objects or classes

(b) *Award [1 max].*
`boolean isFinals;`   **[1]**

Note to examiners: allow any sensible/non-trivial answer, eg. IsDelayed, isCompleted, isHeat, isCancelled, isRecordTime, IsFinished

(c) *Award [3max]*
Award [1] for a generic definition of a class and [1] for a generic definition of an object and [1] for an example that uses either the Swimmer of the Race classes to link together the idea of class and object, for [3 max].

Descriptions
A class is the blueprint/template/abstraction for an entity/object;
An instantiation is the process of creating an object / is an actual object (that contains data);

Example involving actual data that links the 2 concepts
The class `Race` stores up to 8 different instantiated objects of the `Swimmer` class and their times;
Swimmer is the generic class for a swimmer in the event. An instantiation of this class would be a specific swimmer with name, school etc.;   **[3]**

(d) (i) *Award [2 max].*
Award [1] for identifying an advantage and [1] for an elaboration of the advantage up to [2 max].

Aggregation allows code reuse / reduces amount of coding;
As the code for the Swimmer object already exists;

Aggregation allows for the better organization of objects;
As the variables/methods of the aggregated data will be kept in the class in which it belongs;   **[2]**

(ii) *Award [2 max].*
Award [1] for identifying a disadvantage and [1] for an elaboration of the disadvantage up to [2 max].

Aggregation increases dependencies;
A change in the `Swimmer` class could have an unexpected consequence for the `Race` class;

Using aggregation can lead to more complex code;
When accessing functions of the aggregated object;   **[2]**

(e)   *Award **[4 max]**.*
*Award **[1]** for correct signature including parameters.*
*Award **[1]** for assigning name and school (don't need* "this" *if parameters have different names).*
*Award **[1]** for a correct loop.*
*Award **[1]** for assigning "empty", " " or* null *to* eventID[i].
*Award **[1]** for assigning* 0 *(or* 0.0*) to* time[i].   **[4]**

Example answer:
```
public Swimmer(String name, String school)
{
  this.name = name;
  this.school = school;
  for(int i = 0, i < 5; i++) // allow use of length
  {
    eventID[i] = "empty"; // allow null
    time[i] = 0;          // allow 0.0
  }
}
```

(f)   *Award **[3 max]**.*
Modern programming languages use Unicode / UTF family to encode characters;
Which uses more bits than ASCII (for each character) / uses 16+ bits per character (approx. 65000+) / extends the range of ASCII;
This allows the use of non-Latin languages / alphabets / character sets (allow example)   **[3]**

15. (a)   *Award **[3 max]**.*
       Each method is defined within its own class.
       Each method is called within an object of that class.
       Therefore the compiler (allow program) knows which method to use.   **[3]**

   (b)   *Award **[4 max]**.*
       *Mark as **[2]** and **[2]**.*

       It promotes code reuse / reduces redundancy;
       Because the parent object holds common data and actions / which speeds up
       development time;

       It reduces maintenance overhead;
       Because you only have to update the parent;

       Can speed up development time;
       As parent code will (probably) already have been tested;

       The extensibility feature of inheritance;
       Is a measure of the ability to extend the class without having to create completely
       new classes;   **[4]**

       Do not accept generic statements such as "saves time".

   (c)   *Award **[2 max]**.*
       *Award **[1]** for a clear understanding of method overriding and **[1]** for relating it to
       this situation up to **[2 max]**.*

       Method overriding redefines/replaces/overrides a method from the inherited
       class;
       The constructor could only instantiate the `finals` object;
       The method `addSwimmers()` could fill `finals` directly;
       The method `fillFinals()` could do nothing;   **[2]**

16.   (a)   *Award **[3 max]**.*
*Award **[1]** for array of 8 slots with array name (allow any format as long as it is clear).*
*Award **[1]** for correct entries (in any order). Ignore inclusion of times.*
*Award **[1]** for correct order.*

swimmer	Bella	Andy	Eric	Idan	Duc	George	Karl	Hetty
	[0]	[1]	[2]	[3]	[4]	[5]	[6]	[7]

Do not penalize the lack of subscripts or the inclusion of times   **[3]**

(b)   *Award **[6 max]**.*
*Award **[1]** for correctly declaring 2 arrays of size 16.*
Either (Example 1)
   *Award **[1]** for correct outer loop. // allow length function for loops*
   *Award **[1]** for correct inner loop.*
Or (Example 2)
   *Award **[2]** for two correct single loops.*
Or (Examples 3 or 4)
   *Award **[2]** for one loop*
*Award **[1]** for assignment of swimmer object (with accessor)*
*Award **[1]** for assignment of time (with accessor)*
Allow 1 mark if accessors omitted from above 2 answers but otherwise correct
*Award **[1]** for first initialising and then incrementing the index of the new arrays.*

Example answer 1:

```
Swimmer[] tempSwim = new Swimmer[16];
double[] tempTime = new double[16];
int newIndex = 0;
for(int i = 0; i < 2; i++)
    for(int j = 0; j < 8; j++)
    {
        tempSwim[newIndex] = races[i].getSwimmer(j);
                                       // allow … getSwimmer()[j]
        tempTime[newIndex] = races[i].getTime(j);
                                       // allow … getTime() [j]
        newIndex++;
    }
}
```

Note to examiners – the alternatives shown above apply to every example

Example answer 2:

```
Swimmer[] tempSwim = new Swimmer[16];
double[] tempTime = new double[16];
int newIndex = 0;
for(int j = 0; j < 8; j++)    // allow j<5
{
    tempSwim[newIndex] = races[0].getSwimmer(j);
    tempTime[newIndex] = races[0].getTime(j);
    newIndex++;
}
for(int j = 0; j < 8; j++)     // allow j<6
{
    tempSwim.[newIndex] = races[1].getSwimmer(j);
    tempTime.[newIndex] = races[1].getTime(j);
    newIndex++;
}
```
Note: if 2nd loop is incorrect you can still award marks for correct assignments in 1st loop

Example answer 3:

```
Swimmer[] tempSwim = new Swimmer[16];
double[] tempTime = new double[16];
for(int j = 0; j < 8; j++)
{
    tempSwim[j] = races[0].getSwimmer(j);
    tempTime[j] = races[0].getTime(j);
    tempSwim[j+8] = races[1].getSwimmer(j);
    tempTime[j+8] = races[1].getTime(j);
}
```

Example answer 4:

```
Swimmer[] tempSwim = new Swimmer[16];
double[] tempTime = new double[16];
for(int j = 0; j < 16; j++)
{
    if (j < 8)
    {
        tempSwim[j] = races[0].getSwimmer(j);
        tempTime[j] = races[0].getTime(j);
    }
    else
    {
        tempSwim[j] = races[1].getSwimmer(j-8);
        tempTime[j] = races[1].getTime(j-8);
    }
}
```

[6]

Note to examiners: that the question asks for all entries to be copied. However, do not penalize "efficient" solutions that avoid copying the null and 0 entries.

(c) (i)   Bubblesort;   **[1]**

     (ii) *Award **[4 max]**.*
           *Award **[1]** for stating an improvement and **[1]** for an elaboration up to*
           ***[2 max]**.*
           *Mark as **[2]** and **[2]**.*

           Include a flag "swapped";
           That can help stop the outer loop if there is a pass through the inner loop
           with no swap;

           Limit the inner loop by deducting the outer loop counter;
           So that the sorted elements are no longer compared;   **[4]**

(d) *Award **[6 max]**.*
     *Award **[1]** for correctly initialising variables k and count*
     *Award **[1]** for correct loop **

     *Award **[1]** for checking for 0 time (or null student name)*
     *Award **[1]** for correct object (e.g. free100 or finals) **and** correct method*
     *Award **[1]** for passing the parameter* tempSwimmer[k]
     *Award **[1]** for passing the parameter* count
     *Award **[1]** for correct incrementation of variables (and exiting loop in example 4)*

      * *Note: If the student only looked at the first 8 entries and hasn't realised that there are null*
     *entries, award the loop mark **[1]** for a loop of **for (int i=0; i < 8; i++)***

     Example 1 using the undefined mutator method setSwimmer() in the Race class

```
int k=0;
int count = 0;
while ((count<8) && (k<16))      // allow if (k<16) missing
{
     if (tempTime[k] != 0)        // tempSwimmer[k].getName() != null
                                   // allow tempSwimmer[k] != null
     {
          finals.setSwimmer(tempSwimmer[k],count);
          count ++;
     }
     k++;
}
```

     Example answer 2 using the undefined fillFinals() method the Event class:

```
int k =  0;
int count = 0;
while ((count < 8) && (k < 16)) // allow if (k<16) is missing
{
   if (tempTime[k] != 0);            // tempSwimmer[k].getName() != null
                                      // allow tempSwimmer[k] != null
    {
       free100.fillFinals(tempSwimmer[k],count);
       count++;
    }
    k++;
}
```

Example answer 3 using the undefined addSwimmers() method in the Race class:

```
int k =  0;
int count = 0;
while ((count < 8) && (k < 16)) // allow even if (k<16) is missing
{
  if (tempTime[k] != 0)               // tempSwimmer[k].getName() != null
                                      // allow tempSwimmer[k] != null
    {
        finals.addSwimmers(tempSwimmer[k],count);
        count++;
    }
    k++;
}
```

Example 4 using a single for loop

Note that all 3 above versions can be written with a for loop as shown below (the same mark scheme can be used):

```
int count = 0;
for (int k = 0; k < 16; k++)
{
      if (tempTime[k] != 0)     // tempSwimmer[k].getName() != null
                                // allow tempSwimmer[k] != null
      {
          finals.setSwimmer(tempSwimmer[k],count);
          count ++;
          if (count == 8)
              break;
      }
}
```

[6]

17. (a) *Award [1 max].*
A reference is a variable whose value points to the location of an object (in memory); **[1]**

(b) *Award [2 max].*
the ease of which pointers or references can be manipulated in a linked list;
Allows easier addition/deletion of objects (compared to a binary tree);

It can be difficult to manipulate pointers or references in a binary tree;
Which makes it more difficult to add or delete objects in a binary tree; **[2]**

(c) (i) *Award [1 max]. Allow missing final parentheses.*
```
LinkedList<Visitor> guests = new LinkedList<Visitor>();
```
[1]

(ii) *Award [4 max].*
Award [1] for correct signature.
Award [1] for using size() or finding the correct size of the linked list *
Award [1] for correctly returning the penultimate element.
Award [1] for correctly returning null if there is no penultimate element in the list. **[4]**

```
public Visitor penultimate(LinkedList guests) // parameter
is optional
{
  Visitor result = null;
  if(guests.size() > 1) {
    result = guests(guests.size()-2);
  }
  return result;
}
```

*Note to examiners – allow a method that will iterate through the class as long as a valid LinkedList class methods are used.
See https://docs.oracle.com/javase/7/docs/api/java/util/LinkedList.html

(d)   *Award [4 max].*
       Award [1] for showing the second call to the method
       Award [1] for showing the third call to the method (and no subsequent calls)
       Award [1] for including the correct test outcome
       Award [1] for correct output

```
recursive(0,'F')
   k = 0
   current is Ana
   test incorrect
   recursive(1,'F')
        k = 1
        current is Ben
        test incorrect
        recursive(2,'F')
          k = 2
          current is Defne
          test correct
          output Defne Tasci
```

(e)   *Award [2 max].*
       For a large linked list, this would require a large number of recursive calls /
       become very memory intensive;
       Which may cause stack overflow;

       A linked list only allows for sequential access;
       Therefore, using recursion would not lead to any gains in efficiency;   **[2 marks]**

(f)   *Award [6 max].*
        Award [1] for the correct initialisation of index.
        Award [1] for correct loop/iterator declaration and initialisation
        Award [1] for the correct assignment to current.
        Award [1] for the correct comparison of school (allow use of =)
        Award [1] for the correct remove (allow correct use of next pointer)
        Award [1] for the decrementing index after remove.
        Award [1] for correct increment of index / use of `iterator.next()`. Within
        context.   **[6 max]**

        Example answers:

```java
public void removeSchool(String school) {
   int index = 0;
   Visitor current;
   while(index < guests.size()) {
     current = guests.get(index);
     if(current.getSchool().equals(school)) {
       guests.remove(index);
                index--;
     }
     index++;
   }
}


public void removeSchool(String school) {

        Visitor current;
        for(int index = 0; index < guests.size();index++) {
            current = guests.get(index);
            if(current.getSchool().equals(school)) {
                guests.remove(index);
                index--;
            }
        }
}


public void removeSchool(String school) {
   Iterator<Visitor> itr = guests.iterator();
   Visitor current;
   while(itr.hasNext()){
     Visitor current = itr.next();
     if(current.getSchool().equals(school)) {
       itr.remove();
     }
   }
}
```