

**COMPUTER SCIENCE TRIPOS Part IA 75%, Part IB 50%**

---

Thursday 4 June 2020 1.30 to 4.30

---

## COMPUTER SCIENCE Paper 3

Answer **one** question from each of Sections A, B and C, and **two** questions from Section D.

Submit the answers in five **separate** bundles, each with its own cover sheet. On each cover sheet, write the numbers of **all** attempted questions, and circle the number of the question attached.

**You may not start to read the questions  
printed on the subsequent pages of this  
question paper until instructed that you  
may do so by the Invigilator**

## STATIONERY REQUIREMENTS

*Script paper**Blue cover sheets**Tags*

## SPECIAL REQUIREMENTS

*Approved calculator permitted*

## SECTION A

## 1 Databases

Suppose we have a relational database with five tables.

table	key attributes
S(sid, A)	sid
T(tid, B)	tid
U(uid, C)	uid
R(sid, tid, D)	sid, tid
Q(tid, uid, E)	tid, uid

Here R implements a many-to-many relationship between the entities implemented with tables S and T, and Q implements a many-to-many relationship between the entities implemented with tables T and U.

- (a) Write an SQL query that returns all records of the form `sid, uid` where `sid` is the key of an S-record and `uid` is the key of a U-record and these two records are related through the relations R and Q. Use `SELECT` and not `SELECT DISTINCT` in your query. [3 marks]
- (b) Write an SQL query that returns records of the form `A, C` where the A-value is from an S-record and the C-value is from a U-record and these two records are related through the relations R and Q. Use `SELECT` and not `SELECT DISTINCT` in your query. [3 marks]
- (c) Could one of your queries from parts (a) and (b) return more records than the other? If so, which one? Justify your answer. [3 marks]
- (d) Suppose you replaced `SELECT` with `SELECT DISTINCT` in your queries from parts (a) and (b). Could one of these modified queries return more records than the other? If so, which one? Justify your answer. [3 marks]
- (e) Consider again your query from part (a). If pair `sid, uid` is returned by this query then there must exist at least one “path” that goes from from table S to table T (via relation R) and then from table T to table U (via relation Q). Note that there can be many such paths for a given pair `sid, uid`. Write an SQL query that returns records of the form `tid, total` where `tid` is a key of a record from table T and `total` indicates the total number of such paths that “go through” that record. [8 marks]

## 2 Databases

Suppose we have a relational database with three tables

table	key attributes
S(sid, A)	sid
T(tid, B)	tid
R(sid, tid, C)	sid, tid, C

The following referential integrity constraints hold on the table R. Every `sid`-value in R is the key of a record in table S and `tid`-value in R is the key of a record in table T.

- (a) What does it mean that the attribute C is included in R's key? [2 marks]
- (b) Write an SQL query to return records of the form `sid` where `sid` is the key of an S record that is not R-related to any records in table T. [2 marks]
- (c) Consider the following two queries.

```
Q1 = SELECT S1.A AS A, R1.C AS C
      FROM R AS R1
      JOIN R AS R2 ON R2.tid = R1.tid
      JOIN S AS S1 ON S1.sid = R1.sid
      JOIN S AS S2 ON S2.sid = R2.sid
```

and

```
Q2 = SELECT DISTINCT S1.A AS A, R1.C AS C
      FROM R AS R1
      JOIN R AS R2 ON R2.tid = R1.tid
      JOIN S AS S1 ON S1.sid = R1.sid
      JOIN S AS S2 ON S2.sid = R2.sid
```

Note that the only difference is the use of `DISTINCT` in Q2.

- (i) If Q1 and Q2 return the same results, what can you conclude about the data in this database? Justify your answer. [4 marks]
- (ii) Suppose we add this where-clause to each query:

```
WHERE R1.C = R2.C AND S1.sid <> S2.sid
```

If the modified Q1 and Q2 return the same results, what can you conclude about the data in this database? Justify your answer. [4 marks]

- (iii) Suppose we add the where-clause `WHERE R1.tid <> R2.tid` to each query. If the modified Q1 and Q2 return the same results, what can you conclude about the data in this database? Justify your answer. [4 marks]
- (iv) Suppose we add this where-clause to each query:

```
WHERE R1.tid <> R2.tid AND S1.sid <> S2.sid
```

If the modified Q1 and Q2 return the same results, what can you conclude about the data in this database? Justify your answer. [4 marks]

## SECTION B

### 3 Introduction to Graphics

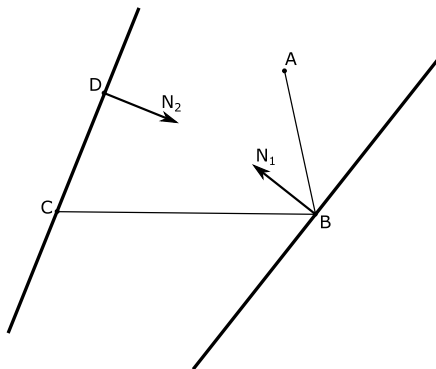
- (a) What is the purpose of mapping from scene-referred to display-referred colours? [3 marks]
- (b) How does display encoding differ between standard- and high-dynamic-range displays? [3 marks]
- (c) Explain how simulating glare that happens in the eyes or in a camera can enhance the appearance of rendered scenes. [3 marks]
- (d) Why is glare typically simulated only for the values that exceed the maximum displayable value? [3 marks]
- (e) Write the equation of a function that maps linear input colour channel value  $x$  (arbitrary range) to display-encoded pixel value such that: (i) the linear value  $x_{white}$  is mapped to the peak value of the display; (ii) the resulting pixel value is in the range from 0 to 255; (iii) the colour is display encoded using the gamma of 2.2. [3 marks]
- (f) You are given spectra of two colours:  $c_1(\lambda)$ ,  $c_2(\lambda)$  and cone response functions  $l(\lambda)$ ,  $m(\lambda)$ ,  $s(\lambda)$ . How would you test whether the colours  $c_1$  and  $c_2$  are metamers? Write down the equations for such a test. [5 marks]

#### 4 Introduction to Graphics

(a) Phong's reflection model is given by the equation

$$I = I_a k_a + \sum_i I_i k_d (\mathbf{L}_i \cdot \mathbf{N}) + \sum_i I_i k_s (\mathbf{R}_i \cdot \mathbf{V})^n$$

- (i) Draw a schematic diagram showing a flat surface, two point-light sources, a camera, a ray-intersection point on the surface and all relevant vectors. Label each vector using the symbols from the equation above. Use indices 1 and 2 for the point lights. [2 marks]
- (ii) Explain what type of reflection is modelled by the *diffuse* term and what parameter(s) control it. [2 marks]
- (iii) Explain what type of reflection is modelled by the *specular* term and what parameter(s) control it. [2 marks]
- (iv) Explain what type of reflection is modelled by the *ambient* term and what parameter(s) control it. [2 marks]
- (v) What does the dot product  $\mathbf{L}_i \cdot \mathbf{N}$  represent and what physical property does it model? [3 marks]
- (vi) How does the distribution of microfacets differ on the surface of diffuse, imperfect specular and perfect specular materials? [3 marks]
- (b) A ray originates at point  $A$ , hits a perfectly specular plane with normal  $N_1$  at point  $B$ , gets reflected and hits point  $C$  on the plane defined by normal  $N_2$  and point  $D$ . Refer to the figure below. Use vector algebra to find the coordinates of point  $C$  given all the other points and vectors listed in the figure. [6 marks]



*This page is intentionally left blank.*

## SECTION C

### 5 Interaction Design

- (a) During your practical session, you were asked to create a working Weather App for a chosen primary stakeholder that worked on a desktop or a laptop. This was done in three stages: (i) gathering requirements and data, (ii) designing a lo-fi prototype, and (iii) implementing a hi-fi prototype.

CUSTOM stakeholder analysis is one of the approaches used for gathering and analysing requirements. Describe what requirements gathering and analysis is and list its main aims. Describe the four different kinds of stakeholder in a CUSTOM analysis, providing for each an example stakeholder from the Weather App your group has developed. [6 marks]

- (b) Describe three limitations of task analysis. Explain whether any of these limitations would affect the analyses of the tasks for the Weather App developed during your practical session or not, and why. [4 marks]

- (c) Consider each of the findings below and the associated summary statement about it. For each one, comment on whether the finding supports the statement.

(i) Finding: Two out of four people who filled in the questionnaire ticked the box that said they prefer not to use the ring-back facility on their cell phone. Statement: Half of the users do not use the ring-back facility.

(ii) Finding: Joan who works in the design department was observed one day walking for 10 minutes to collect printout from the high-quality colour printer. Statement: Significant time is wasted by designers who have to walk a long distance to collect printout.

(iii) Finding: A data log of 1000 hours of interaction with a website recorded during January, February, and March records 8 hours spent looking at the helpfiles. Statement: The website's helpfiles were used less than 1% of the time during the first quarter.

[6 marks]

- (d) A company has been asked to design and implement a system for visually impaired people. The main requirement is that the system will work with a lightweight device that the user can wear on her forehead, and the device will allow the user to hear colour by turning colour into sound frequencies. So, instead of seeing the world in grayscale, the user will be able to hear a symphony of colour, and even listen to faces and paintings. Consider how you would use Heuristic Evaluation and Cognitive Walkthrough techniques to evaluate this system. How would the two evaluation techniques compare for evaluating this system and why? [4 marks]

## 6 Interaction Design

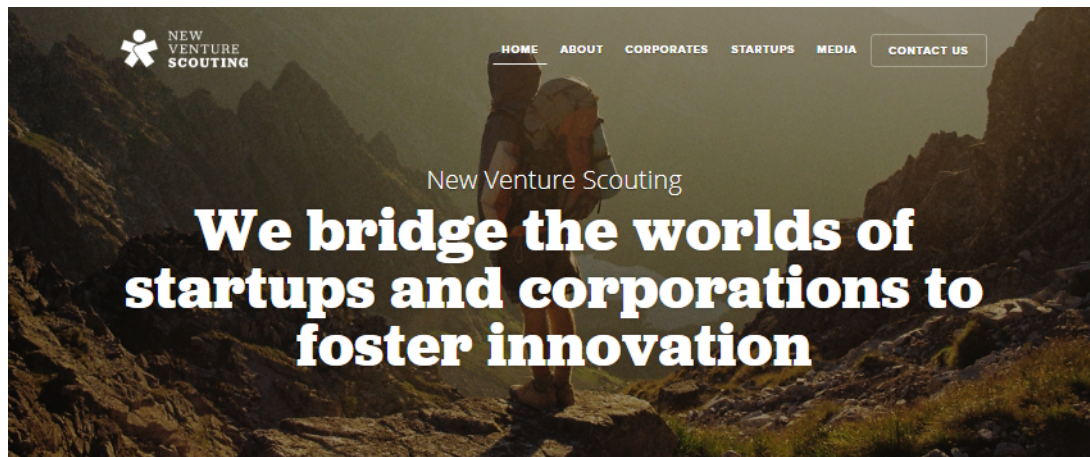
- (a) During your practical session you were asked to create a working Weather App for a group of chosen primary stakeholders that worked on a desktop or a laptop. This was done in three stages: (i) gathering requirements and data, (ii) designing a lo-fi prototype, and (iii) implementing a hi-fi prototype.

Describe the primary stakeholders the Weather App your group has developed for, and the three tasks your App supports for the chosen primary stakeholders together with the justification for each task supported. Consider the task of 'Find today's temperature in city X' and sketch a task model for this assuming that the Weather App is not yet open. [6 marks]

- (b) A company has been asked to design and implement a system for visually impaired people. The main requirement is that the system will work with a lightweight device that the user can wear on his forehead, and the device will allow the user to hear colour by turning colour into sound frequencies. So, instead of seeing the world in grayscale, the user will be able to hear a symphony of colour, and even listen to faces and paintings. Explain which data collection techniques would be appropriate and which ones would be difficult to use for gathering requirements, and describe why that is the case. [4 marks]

- (c) Provide a definition of Gestalt theory in the context of interaction design. List the six principles of Gestalt psychology that have implications on interaction design, explaining what each principle is about. Describe how each of these principles were employed in the design of the web page given in the figure below. [10 marks]





### Discover exciting startups

Get access to the ideas of tomorrow

Startups are the "Laboratories of the future". Their flexibility, proximity to the customer and lean structures are key assets in our complex business world. NVS opens the door to getting to know the right startups that complement your business.



### Bring innovation into your company

Working with startups improves your innovation process

Markets are changing faster than companies and their business models. There's a high pressure to innovate, but in-house innovation processes need time and they are expensive. Cooperating with startups gives you a head start on products and business models of the future.



### Dive into entrepreneurial culture

Bring new impulses to your company culture

The best results can always be expected when a diverse group of people -your team and dynamic entrepreneurs- are working together. Therefore, getting in touch with the vibrant startup culture that has taken up speed all over the world will challenge outdated assumptions.

## How we work in 4 steps

We value a targeted approach with the aim of bringing innovation into your company and connecting startups and corporations on eye-level



Defining innovation needs



Global startup scouting



Screening & Filtering



Facilitating cooperation

- [HOME](#)
- [ABOUT](#)
- [CONTACT US](#)
- [IMPRINT](#)
- [DATA PRIVACY POLICY](#)

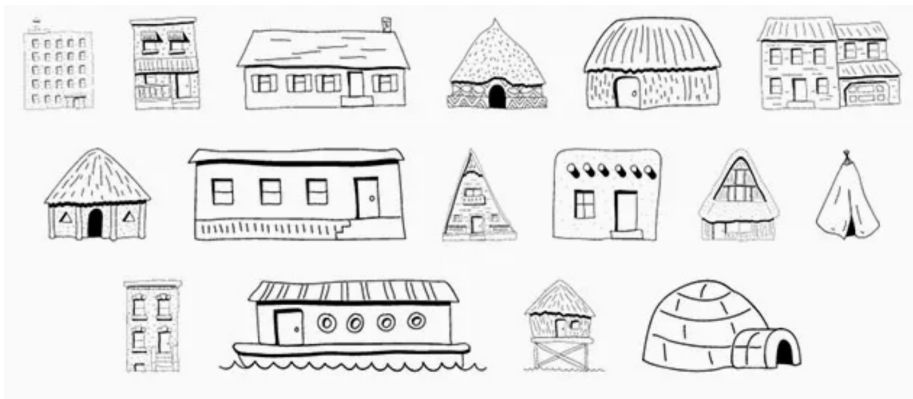
**CONTACT US**  
 email: [office@nvs.co.at](mailto:office@nvs.co.at)  
 phone: +43 899 199 83 000

**ADDRESS**  
 St. Paul GmbH  
 Breitenfurter Straße 25/21  
 1120 Vienna  
 Austria  
[View on Google Maps](#)

## SECTION D

## 7 Machine Learning and Real-world Data

You are part of a team of developers whose task is to classify hand-drawn images into classes defined by the objects they depict, in the framework of a new search engine that is based on shapes. The images are in black and white and are represented by  $150 \times 200$  pixel bitmaps. Consider the figure below, which shows some examples of such images for the classes “house”, “tent”, “boat” and “igloo”.



- (a) You want to use a Naive Bayes Classifier with a “bag of pixels” representation.
- (i) Describe which approach you would use for classification, including which features you would use and how many there are. State the relevant formula. [5 marks]
- (ii) What does the training material for the task look like, how was it gathered from human annotators, and how is it used to estimate the model’s parameters? [4 marks]
- (b) To improve classification, you plan to preprocess the raw pixel features in some form so that a smaller, more informative set of features results. Describe the best such manipulation you can think of, and why this should improve results. [4 marks]
- (c) Your colleague suggests that instead of pixel-based features, you should derive some higher-level features for the classifier, which capture holistic visual properties of the image instead of exact pixel location.
- (i) Describe three such features which you think will be most helpful in classification, giving reasons why this feature should work and using the above figure for illustration. Choose maximally different features. [3 marks]
- (ii) Describe how the final classifier can be modified so that it combines pixel-based features with your features from Part (c) above. What possible problems do you foresee, and how could they be addressed? [4 marks]

## 8 Machine Learning and Real-world Data

A farmer has been collecting data about the performance of a crop. Every year, one of three methods is used for cultivation: fertilizer (F), pesticide (P) or none (N). The farmer characterizes the performance of the crop as good (G), average (A) or bad (B). The following data has been collected:

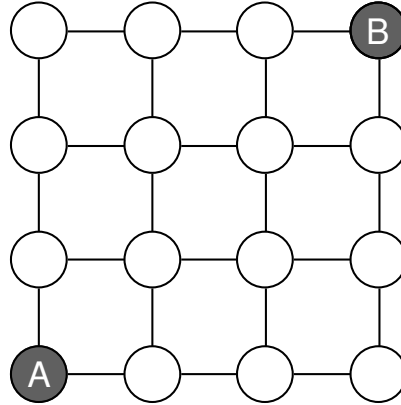
```
F P N F F N P P N N F P
A A B G A B G A B A G B
```

The farmer wants to apply a Hidden Markov Model (HMM) to understand the relation between cultivation methods and crop performance, treating the cultivation method as the hidden states and the crop performance as the observations.

- (a) Describe the components of the first order HMM using equations where appropriate. For the parameters you define, also describe how you would estimate them from the data given above. [7 marks]
- (b) What are the assumptions made in the first order HMM? What do they mean in the context of the farming application described above? [4 marks]
- (c) Given the HMM and the farming data above, answer the following questions, explaining your answers.
- (i) Which cultivation method is more likely to result in a good crop?
- (ii) Which succession of cultivation methods is more likely to result in a succession of two good crop years? Assume that for the first of the two years the choice of cultivation method is uniform.
- (iii) Is a bad crop year more likely to be followed by an average or a good crop year? [6 marks]
- (d) By this point you have solved the labelled learning problem for the HMM. Name the other three problems described below: [3 marks]
- (i) Given the observations about crop performances and knowledge of the possible cultivation options, find the parameters of the HMM.
- (ii) Given the parameters of the HMM and a sequence of crop performances, determine how likely the sequence of performances is.
- (iii) Given the parameters of the HMM and a sequence of crop performances, infer the most likely sequence of cultivation methods used.

## 9 Machine Learning and Real-world Data

Consider the following network, where the nodes are arranged in a four-by-four grid. Each node is only connected to its neighbours. We distinguish two nodes: The bottom-left-hand-corner node, which we call A, and the upper-right-hand-corner, which we call B.



- (a) Shortest paths.
- (i) What is the definition of a shortest path? [1 mark]
  - (ii) How many shortest paths are there from node A to node B? [2 marks]
- (b) Diameter of a graph.
- (i) What is the definition of the diameter of the graph? [1 mark]
  - (ii) What is the diameter of the graph presented above? [2 marks]
- (c) Betweenness centrality.
- (i) What is the definition of betweenness centrality? (Please give the formula.) [2 marks]
  - (ii) Name a node with minimum betweenness centrality in the graph presented above. [3 marks]
- (d) Local bridges.
- (i) What is the definition of a bridge? [1 mark]
  - (ii) What is the definition of a local bridge? [1 mark]
  - (iii) Is there a bridge in the graph presented above? If so, name the edge. If not, explain why not. [2 marks]
  - (iv) Is there a local bridge in the graph presented above? If so, name the edge. If not, explain why not. [2 marks]
- (e) Do real-world networks look like the one in the figure above? If so, describe one. If not, name the properties of the network that make it unrealistic. Be clear and precise using terminology from the lecture. [3 marks]

**END OF PAPER**