CST0
COMPUTER SCIENCE TRIPOS Part IA

Friday 13 June 2025   13:30 to 16:30

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

(a) X, Y and Z are sets stored as relations. A user wishes to compute $X \cap (Y \cup Z)$ and writes the following query:

```
select X.A from X join Y join Z where X.A=Y.A or X.A=Z.A
```

Explain whether this SQL is correct or provide a working SQL query.

[4 marks]

(b) This part is about lists and sets stored as relations. Your answers must not use modern features of SQL that allow collections inside rDBMS fields.

(i) Give two differences between a set and a list in Computer Science.

[1 mark]

(ii) Give an rDBMS schema for a list. Give also a schema for a set. Each schema should consist of just one table. Briefly explain why your definitions are correct. [3 marks]

(iii) Consider the following two schemas: can either precisely represent a list of sets or a set of lists as one relation? Could one represent a list of lists?

[5 marks]

$$\text{Foo}(\underline{N, T}, A) \qquad \text{Bar}(\underline{N, T, A})$$

A is an item, N is a natural number and T is nominally the name of a set.

(c) (i) For three-valued logic, give a real-world example where a non-commutative outer join (ie. left or right-outer join) is useful. [1 mark]

(ii) Using the Relational Algebra or basic set theory, quote a baseline definition of the equijoin operation $R \bowtie_= S$ for two-valued logic (ie. where null values are not present). [2 marks]

(iii) Provide a similar definition for the left-outer join for three-valued logic. *Hint: your definition might use a tuple containing only null values, $(\omega, \ldots, \omega)$, of length equal to the number of attributes in R that are not in S.* [4 marks]

2

2 Databases

(*a*) A vineyard makes several types of wine that use varying ratios of up to three different varieties of grape and are shipped in various shapes of wine bottle.

    (*i*) Draw a suitable E/R diagram, including a few additional attributes that are likely in the real world. [1 mark]

    (*ii*) Give a schema for an rDBMS that hardcodes the limit of three grape types and suggest at least two rules of consistency. [3 marks]

    (*iii*) Give a second rDBMS schema, where there is no limit to the number of grape types in a wine. [2 marks]

    (*iv*) Using a graph database, give a schema or concrete example for the unrestricted case. [2 marks]

    (*v*) Briefly discuss the relative merits of the three schemas. [2 marks]

(*b*) A friend has typed out a collection of old recipe cards and stored them in an rDBMS table with two fields: (recipe_name, recipe_text). I have suggested he could instead split out the ingredients part of each recipe, storing it in structured JSON inside a third field.

    (*i*) Using a traditional rDBMS, discuss these two approaches. [3 marks]

    (*ii*) A new rDBMS provides limited support for tables where fields contain JSON by notionally expanding the JSON into a form where it can be readily queried with traditional SQL. What problems might remain? Precisely describe the operation of two useful JSON expansions. Note that more than one field might contain JSON. How might either expansion help with the recipes? [7 marks]

(TURN OVER)

SECTION B

3 Introduction to Graphics

(a) Explain what a pinhole camera model is (you may draw a diagram if needed), and what the camera origin in ray tracing corresponds to in the pinhole model.

[2 marks]

(b) Assuming a Phong shading model and no recursion for computing indirect illumination or shadows, what is the time complexity of a ray tracing algorithm and why? You should analyze it by considering the necessary loops and their structure.

[5 marks]

(c) Explain the shadow rays. What is the colour of a pixel if it is in shadow, assuming a Phong shading model and no recursion for computing indirect illumination?

[3 marks]

(d) In the Phong shading model, explain all circumstances where the specular and diffuse components can become zero.

[4 marks]

(e) How does the specular highlight differ from diffuse reflection in the Phong shading model? You do not need to write the formula.

[2 marks]

(f) In ray tracing, how would you compute an approximate full spectrum of colours instead of RGB values for each pixel? What is the time complexity in this case?

[4 marks]

4 Introduction to Graphics

(a) Answer the following questions about geometric transformations.

   (i) Given a cylinder defined at the origin with radius 1 and its ends at $(0, 0, 0)$ and $(0, 0, 2)$, how would you transform it to have radius 2 and its ends at $(1, 2, 5)$ and $(4, 6, 5)$ if the order of transformations is set to be scaling, rotation, and translation? You do not need to compute the matrices, but you should explicitly describe what each transformation does.   [6 marks]

   (ii) How would you obtain the transformation for the normal using the above scaling, rotation, and translation? Simplify the transformation if possible.
   [2 marks]

(b) Answer the following questions about the graphics pipeline.

   (i) In rasterization, what information does the Z-buffer store? Why is this information needed, and how was it computed?   [3 marks]

   (ii) In the OpenGL rendering pipeline, can you give 4 examples of vertex attributes and uniforms, respectively?   [4 marks]

   (iii) Explain the main difference between uniform variables and vertex attributes.   [2 marks]

   (iv) What are barycentric coordinates, and where are they needed in the OpenGL rendering pipeline?   [3 marks]

(TURN OVER)

SECTION C

5 **Interaction Design**

Many young adults are starting to explore career opportunities, apply for jobs, or prepare for internships. However, many lack guidance on resume building, interview skills, networking, and job search strategies. To bridge this gap, your company has been tasked with developing a mobile app (for Android and iOS) that helps young adults in this area of their life.

(*a*) Think about the interactive product you are designing, and identify and describe with rationale three key requirements that the application must meet.

[3 marks]

(*b*) Sketch out a design for this app with relevant screens and details (2–4 screens depending on the detail provided on each screen), illustrating how it meets the set of requirements you have listed in (*a*) by providing labels and explanatory captions, and descriptions of the interactive aspects.   [6 marks]

(*c*) Describe what information architecture is and then create a diagram outlining the information architecture of the app you have designed.   [3 marks]

(*d*) Choose a major task that a potential user can perform with the app you have designed in (*b*), and undertake a Cognitive Walkthrough using the screens you have created.   [6 marks]

(*e*) Using the outcome of the CW you have conducted, provide one suggestion for re-designing your app.   [2 marks]
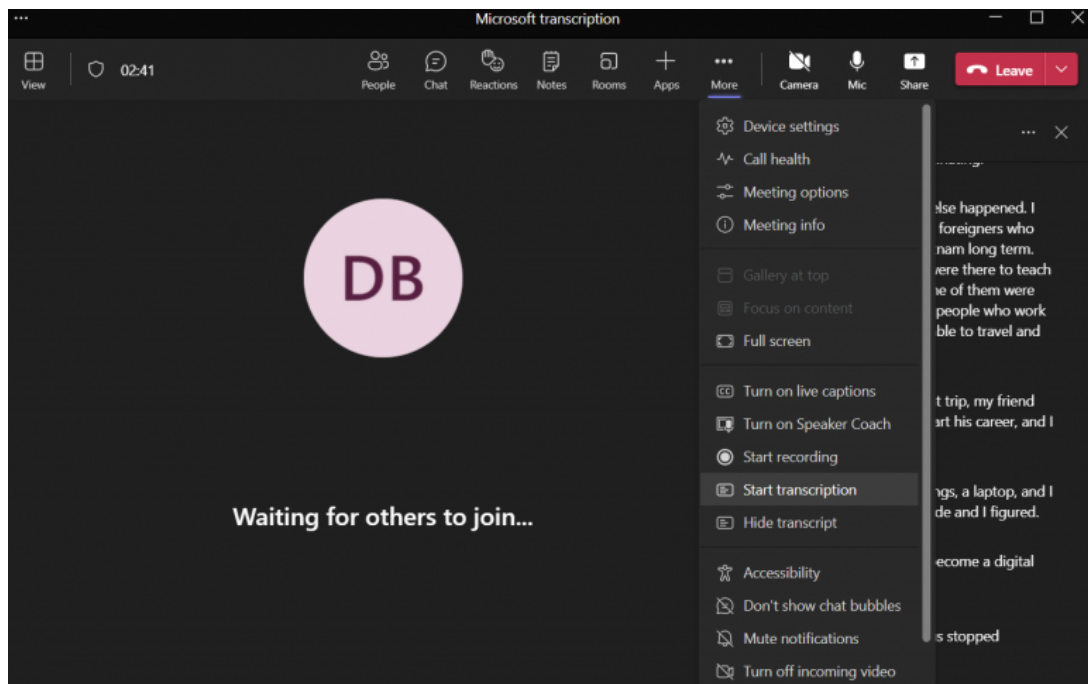
6 Interaction Design

Microsoft Teams is a collaborative workspace within Microsoft 365/Office 365 that acts as a central hub for workplace conversations, collaborative teamwork, video chats, and document sharing, all designed to aid worker productivity in a unified suite of tools. In Microsoft Teams there are several activities that are generally used by users, starting from starting a new team, creating a channel, starting a new conversation, replying to a conversation, starting a meeting, and so on.

(*a*) An evaluator undertook Heuristic Evaluation while evaluating the usability of Microsoft Teams and created the list below. For each problem, identify which heuristic is being violated and briefly explain why.

(*i*) Microsoft Teams can experience buffering or long page loading when there is a slight internet interruption.

(*ii*) There is no demonstration of the use of common tasks.

(*iii*) Sometimes users are confused about starting an activity because there is no wizard that can guide the user.

(*iv*) The layout of objects in Microsoft Teams is not arranged based on the order of steps that need to be taken.

(*v*) Users have difficulty performing common tasks if they were not given training or guidance because the features were numerous and complex.

(*vi*) In Microsoft Teams there is no advanced search feature.

(*vii*) Posts are ranked as most important, namely the last response from the user, such as providing a reply to a form. Meanwhile, files, assignments and other important information are sometimes overwritten by new posts/replies.

(*viii*)The Microsoft Teams search feature in the search engine does not directly display search results that lead to the Microsoft Teams application, but rather "Microsoft" as the parent application.

(*ix*) The navigation in the Microsoft Teams application is located on the left, whereas on the mobile platform it is more difficult to check the navigation.

[9 marks]

(*b*) Considering the figure below and human attention as it applies to interaction design, identify one aspect of human attention and explain how Microsoft Teams made use of this aspect or violated it.

[2 marks]

(c) Considering the figure in (b) and the human memory as it applies to interaction design, identify one aspect of human memory and explain how Microsoft Teams made use of this aspects or violated it. [2 marks]

(d) Considering the Heuristic Evaluation related issues listed in (a), your team was asked to propose a new product-specific design principle that the revised Microsoft Teams app should follow, and to motivate this proposition.

[3 marks]

(e) Your team has been tasked with identifying the reasons behind the low adoption of Microsoft Teams among university students. Describe how you would approach this problem from an interaction design perspective. [4 marks]

SECTION D

7 Machine Learning and Real-world Data

You are tasked with developing a detector of offensive messages using a text classification method. For this purpose a dataset has been provided to you. The dataset contains a sample of 10,000 messages which have been labelled for offensiveness. The files contain information on the text of a message, together with a label ("OK"/"offensive")

(*a*) You want to build your detector using an 'offensiveness' lexicon.

    (*i*) Which steps would you follow to develop such a lexicon? [2 marks]

    (*ii*) How would the detector work? [3 marks]

    (*iii*) How would you evaluate the performance of the detector? Give details of the evaluation metrics and equations, as well as necessary data splits.
    [4 marks]

(*b*) You have the budget to employ 10 human raters to rate 100 texts each. You want to use these ratings to evaluate your system. How would you assess whether the resulting human ratings are reliable for evaluating your system? [3 marks]

(*c*) Explain how you can set up a Naive Bayesian classifier for this task and derive the required parameter estimates.

    (*i*) Give all necessary formulae and discuss smoothing if necessary. [4 marks]

    (*ii*) In what circumstances is the Naive Bayes classifier an improvement over the lexicon-based approach? Explain your answer fully and discuss how we might practically use the lexicon to improve the Naive Bayes approach further. [4 marks]

8 Machine Learning and Real-world Data

Consider the following undirected graph:
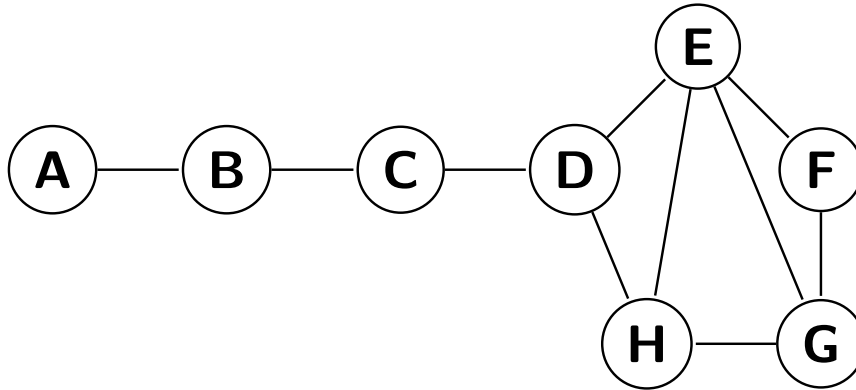
(a) Define the following measures, and compute their values for nodes A, C, and H:

   (i) Degree   [1 mark]

   (ii) Local Clustering Coefficient   [2 marks]

   (iii) Betweenness Centrality   [2 marks]

(b) How could you measure the degree to which nodes form clusters or cliques within a network?   [2 marks]

(c) Which measure would you use to find *connectors* in a social network.
                                                                 [2 marks]

(d) What is the diameter of the network depicted at the beginning of this question?
                                                                 [2 marks]

(e) Suppose that we delete the edge D-H and add an edge C-E to the network. How would this affect the Betweenness Centrality of nodes C, H, and E?   [4 marks]

(f) What is a strongly connected directed graph? Is it possible to change the original graph into a directed graph by adding directionalities to the edges, so that the resulting graph is strongly connected?   [3 marks]

(g) What algorithm could you use to break the original graph into connected clusters? Briefly describe how the algorithm works.   [2 marks]

9 Machine Learning and Real-world Data

Hidden Markov Models (HMMs) can be used for modelling the consonant-vowel sequences in words, using the letters of the words as the hidden states. An HMM of this type has two output options: $q_1 = V$ (for a vowel) and $q_2 = C$ (for a consonant). Each letter in the training data is labelled with either V or C. The training data is as follows (each column below is an example):

t a b l e	t r a c k	c a r	b e t	t a l l	l a b	b e l l
C V C C V	C C V C C	C V C	C V C	C V C C	C V C	C V C C

(a) Provide the full transition matrix A for this HMM based on the training data shown (ignoring the end states). [4 marks]

(b) Give the general formula for calculating emission probabilities from training data, and calculate the emission probabilities for a and ℓ. [2 marks]

(c) Suppose the system is in hidden state a, what is the most likely V, or C, for the next letter, and what are their probabilities. [2 marks]

(d) Suppose the HMM is in hidden state e, what is the most likely V, or C, for the letter 2 letters ahead, and what are their probabilities. [4 marks]

(e) What would be the probability of the following sequence of hidden states?

    b e a t

[2 marks]

(f) How could you improve the model so that unseen combinations of letters do not make a previously unseen word totally impossible? Write the equations and new necessary matrices. What would the probability of

    b e a t

become with this change? [6 marks]

END OF PAPER