COMPUTER SCIENCE TRIPOS Part IB

Tuesday 5 June 2012 1.30 to 4.30

COMPUTER SCIENCE Paper 4

*Answer **five** questions.*

*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

1 Artificial Intelligence I

(*a*) You are designing an *informed search* algorithm to solve a problem of interest. Explain what a *heuristic function* is and why you might want to use one. [2 marks]

(*b*) Define what it means for a heuristic function to be *admissible*, and explain why it might be desirable for such a function to have this property. [3 marks]

(*c*) You have designed two possible heuristic functions h_1 and h_2, both of which you have shown to be admissible, and both of which are applicable to your problem. It has been suggested that you should try to combine them to make a more general heuristic function h defined as

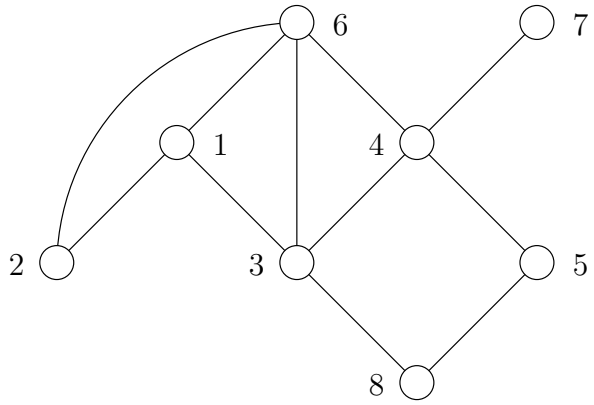
$$h(s) = \alpha_1 h_1(s) + \alpha_2 h_2(s).$$

In this definition s denotes any state in the search problem, and the constants α_1 and α_2 are constrained such that they are non-negative and $\alpha_1 + \alpha_2 = 1$. Is the function h always admissible? Either prove that this is the case or provide a counterexample. [8 marks]

(*d*) For a specific set of states $\{s_1, \ldots, s_n\}$ you have established that the *exact* corresponding distances to the goal are $\{d_1, \ldots, d_n\}$. You want to use this information to make a good choice of the parameters α_1 and α_2 in part (*c*). Derive an algorithm that allows you to do this. You may ignore the constraint requiring the parameters to be non-negative and sum to 1. [7 marks]

2 Artificial Intelligence I

We wish to solve the following graph colouring problem by treating it as a *constraint satisfaction problem*.



The nodes of the graph must be coloured either red (R), blue (B) or green (G) with no pair of connected nodes having the same colour.

(*a*) Describe *Gaschnig's algorithm*. In what way does it improve on *chronological backtracking*? [6 marks]

(*b*) The following sequence of assignments has been made: 1 = R, 2 = G, 3 = B, 4 = G, 5 = B. Explain how Gaschnig's algorithm operates when trying to make an assignment to 6. [3 marks]

(*c*) Describe how *graph-based backjumping* would behave in the situation described in part (*b*). Does it backjump to the same place? Why might graph-based backjumping in general be preferred to Gaschnig's algorithm? [6 marks]

(*d*) Describe how *forward checking* would deal with the sequence of assignments given in part (*b*). How does the effectiveness of backjumping compare with that of forward checking in this case? [5 marks]

3 Computer Graphics and Image Processing

(*a*) What are the main criteria to be considered in the design of a line drawing algorithm for a raster graphics display? [2 marks]

(*b*) Describe an algorithm to fill a series of pixels running from (x_0, y_0) to (x_1, y_1) that meets these criteria, explaining why it does so. Answers should consist of more than a fragment of pseudo-code. [6 marks]

(*c*) A new volumetric display stores an image as a three-dimensional array of volume elements or *voxels*. Reformulate the design and implementation of the line-drawing algorithm to fill a series of voxels running from (x_0, y_0, z_0) to (x_1, y_1, z_1). [6 marks]

(*d*) How would this line-drawing algorithm be used to draw Bézier curves in three dimensions? [6 marks]

4 Computer Graphics and Image Processing

(*a*) Describe how transform coding can be used to compress image data. [4 marks]

(*b*) Explain the Walsh-Hadamard transform on a one-dimensional array of 4 greyscale values. [4 marks]

(*c*) Extend this to the Walsh transform for a two-dimensional array of 2×2 greyscale values. [4 marks]

(*d*) Develop an analogous transform for a three-dimensional array of $2 \times 2 \times 2$ greyscale values. [8 marks]

Include sufficient algebra in your answers to allow a competent programmer to implement the algorithm.

5 Databases

This question explores *Heath's Rule*, which states that if $R(X,\ Y,\ Z)$ satisfies the functional dependency $X \to Y$, where X, Y, and Z are disjoint non-empty sets of attributes, then

$$R = \pi_{X,Y}(R) \bowtie_X \pi_{X,Z}(R),$$

where $\bowtie_X$ is the natural join on the attributes of X.

(a) What is meant by the functional dependency $X \to Y$? [2 marks]

(b) Define the natural join operation $\bowtie_X$. [2 marks]

(c) Suppose that the functional dependency $X \to Y$ holds and we use Heath's rule to justify replacing the schema $R(X,\ Y,\ Z)$ with $R_1(X,\ Y)$ and $R_2(X,\ Z)$.

(i) Give two possible advantages for this schema change. [2 marks]

(ii) Give two possible disadvantages for this schema change. [2 marks]

(iii) Is X a key for R_1? Explain. [2 marks]

(iv) Is X a key for R_2? Explain. [2 marks]

(d) Prove that Heath's Rule always holds. [8 marks]

6 Databases

(*a*) Codd's 1970 paper introduced the *Relational Model* of data to address the difficulties of building database applications using the technology that was available at the time.

(*i*) What problems were encountered by database developers before Codd introduced the Relational Model? [1 mark]

(*ii*) Describe the basic elements of the Model, and explain what is meant by a *relational schema*. [4 marks]

(*iii*) Explain how a formal schema can assist both the application database designer and a database application programmer. What if any are the disadvantages of adopting a mathematical description of database structure? [5 marks]

(*b*) In 1976 Peter Chen introduced the *Entity Relationship (E-R) Model* to support a more natural description of real world data.

(*i*) Describe the basic elements of the Model, and explain some of the choices available to the database designer. [4 marks]

(*ii*) Explain what is meant by a *foreign key* in the relational model. How could you use foreign keys to represent a database described by an E-R model in relational form? To what extent are the two approaches to data modelling complementary? [6 marks]

7 Economics and Law

(*a*) Explain what economists mean by a *lemons market*. [4 marks]

(*b*) What is the difference between *hidden information* and *hidden action*? Give examples of each. [4 marks]

(*c*) Your company is about to launch a new anti-malware product for use on Android devices. To blunt possible competition from low-cost, low-quality vendors you decide you need to offer prospective customers some kind of assurance. Discuss the advantages and disadvantages of the following options.

(*i*) You form a trade association with several other anti-malware firms and promote a quality assurance mark, perhaps with assistance from the government. [4 marks]

(*ii*) You approach the UK banks with a proposal that they certify your product for use along with their banking app products, in the sense that a customer using your system will have exercised due diligence. [4 marks]

(*iii*) You tell customers that if they are the victims of a scam that used malware on their phone you will pay their legal bills. [4 marks]

8 Security I

Briefly explain

(*a*) the function of a *salt value* in a password database [3 marks]

(*b*) *two* examples of covert channels in a file system protocol that is restricted to read-only operations under a mandatory access-control policy [2 marks]

(*c*) *three* types of common software vulnerabilities, with examples [9 marks]

(*d*) *two* problems solved by Cipher Block Chaining [2 marks]

(*e*) under which conditions will user U be able to remove a directory D in Berkeley Unix [4 marks]

9 Security I

(*a*) Name and briefly explain *three* increasingly demanding security properties expected from a *secure hash function* h. [3 marks]

(*b*) Explain how a 128-bit secure hash function h can be used to implement a one-time signature scheme, *and* how to verify such a signature. [5 marks]

(*c*) The T1000 encryption module was designed to make *TeleGroove* messages difficult to read for eavesdroppers. Such messages are character sequences of arbitrary length, encoded using a 5-bit alphabet $A = \{a_0, \ldots, a_{31}\}$. The T1000 module reads blocks of up to 1000 characters at a time into its memory $(m_0, m_1, \ldots, m_{999})$ and then outputs them again in a different order $(m_{K(0)}, m_{K(1)}, \ldots, m_{K(999)})$, according to a secret key table $K : \{0, \ldots, 999\} \to \{0, \ldots, 999\}$ that is shared with the respective recipient of each message. It repeats that process until the block ending with the last character has been processed.

(*i*) What constraint on K ensures that no information is lost? [2 marks]

(*ii*) Identify an ambiguity in the above description and propose a flaw hypothesis regarding a related residual-information vulnerability in this device. [4 marks]

(*iii*) What is the smallest number of blocks that a chosen-plaintext attacker has to send through the device to recover K? Give an example of the content of such blocks and explain how to recover K from them. [6 marks]

END OF PAPER