CST1+CST2
COMPUTER SCIENCE TRIPOS Part IB 75%, Part II 50%

Thursday 10 June 2021 11:30 to 14:30 BST

COMPUTER SCIENCE Paper 7

Answer ***five*** *questions.*

Submit each question answer in a ***separate*** *PDF. As the file name, use your candidate number, paper and question number (e.g.,* `1234A-p7-q6.pdf`*). Also write your candidate number, paper and question number at the start of each PDF.*

You must follow the official form and conduct instructions for this online examination

1 Concepts in Programming Languages

(*a*)
```
static void fastFilter(ArrayList<Datum> a, Predicate<String> p) {
  int j = 0;
  for (int i = 0; i < a.size(); i++) {
    if (p.test(a.get(i).key)) a.set(j++, a.get(i));
  }
  a.removeRange(j, a.size());
}
```

The above method `fastFilter` operating on `ArrayList<Datum>` was written by a Java programmer, perhaps misguidedly in the interests of speed. It uses method `removeRange` to remove the final `a.size()-j` elements from `a`. We wish to enable `fastFilter` also to operate on `ArrayList<DatumDash>` where `DatumDash` inherits from `Datum`.

(*i*) Give code for a modified version using *subtype* polymorphism, justifying any changes and highlighting any unresolvable difficulties. [5 marks]

(*ii*) Similarly, what changes, if any, would be required if the original `fastFilter` instead used Java classic `[]` arrays? Again justify your answer. [3 marks]

(*iii*) Give a definition of `fastFilterInner` which uses *generic* polymorphism, along with a subtype-polymorphic wrapper whose signature matches that used in Part (*a*)(*i*). [*Hint:* consider moving the use of `.key`.] [4 marks]

(*b*)
```
class Myst {                        // A Mystery?
  private Supplier<Int> act;        // Java's name for void->Int
  private Myst(Supplier<Int> a)  { act = a; }
  static Myst R(Int x)          { return new Myst(()->x); }
  Myst B(Function<Int,Myst> f) { return f.apply(act.get()); }
  static Myst primIn = new Myst( () -> SystemIO.readInt()) );
  static Myst primOut(Int x) {
    return new Myst(() -> { SystemIO.println(x); return x;} );
  }
  static void exec(Myst x) { x.act.get(); }
}
```

The `class Myst` is coded in a language resembling Java. What concept does it model and what does the code below conceptually do if executed in its scope?

```
Myst one = primIn.B(x -> primOut(x+1));
Myst two = R(2).B(y -> primIn.B(x -> primOut(x+y)));
Myst three = one.B(x -> two).B(x -> one);
exec(three);
```

[8 marks]

2 Economics, Law and Ethics

The new Master of (fictional) Porterhouse would like to modernise the college. They bring the following proposals to the College Council, of which you are a member.

(*a*) Proposal 1: Change the way students are admitted to the college for all subjects. Rather than having humans interview candidates, or at least read their references and personal statements, an algorithm would make all admission decisions.

(*i*) Use an ethical perspective to make a persuasive argument for or against Proposal 1, defining and explaining relevant concepts. [5 marks]

(*ii*) If Proposal 1 is accepted, what changes would the college need to implement to comply with GDPR? [5 marks]

(*b*) Proposal 2: The college is buying apartments in the next phase of the North West Cambridge development for use by graduate students and students with dependents. The options are to buy a) a range of apartments, from utilitarian to luxurious, which would be rented out at different amounts; or b) a block of apartments all of the same size, which would be rented out at the same price.

(*i*) Use an economic perspective to make a persuasive argument for option a) or b), defining and explaining relevant concepts. [5 marks]

(*ii*) If Proposal 2, option a) is accepted, how might the college decide how much to charge for each grade of apartment? [5 marks]

3 Economics, Law and Ethics

(*a*) You incorporate existing open source code into a program you are writing. What intellectual property aspects will you need to consider? [4 marks]

(*b*) What are the economic arguments for and against making code you write open source? [6 marks]

(*c*) You are approached by someone who claims to have identified a vulnerability in your popular software application. They demand you make them a payment, otherwise they will release the vulnerability publicly. The vulnerability is fixable and has not (yet) been exploited, however it would likely generate adverse publicity. What is the ethical approach to take? [5 marks]

(*d*) You are approached by someone who claims to have identified a vulnerability in your corporate servers and breached your customer database. They demand you make them a payment, otherwise they will release the personal information. Your logs confirm the database has been breached. What is the lawful approach to take? [5 marks]

4 Formal Models of Language

(*a*) The following is a pattern for some legal strings in a language:

$$\texttt{[}a \in A\texttt{]\{0,1\}}\ \ \texttt{[}b \in B\texttt{]\{0,1\}}\ \ \texttt{[}c \in C\texttt{]\{1,n\}}\ \ \texttt{[}d \in D\texttt{]\{1,1\}}$$

where A is a finite set of characters from the alphabet, Σ; similarly for B, C, D. The sets A, B, C and D are disjoint. $\{x, y\}$ indicates that the previous bracket must match at least x times but no more than y times.

(*i*) Specify a Deterministic Finite Automaton, M_1, that can recognise these strings only. [4 marks]

(*ii*) Design a Regular Grammar, G_1, which generates $L(M_1)$. [4 marks]

(*iii*) Describe a set of strings in a natural language that could be generated by G_1 given an appropriate Σ and its subsets A, B, C and D. [1 mark]

(*b*) We can hypothesise that matches of the following pattern are always valid constructions in English:

$$\texttt{[The }Noun\texttt{]\{n,n\}}\ \ \texttt{[}Verb\texttt{]\{n,n\}}$$

where $Noun$ represents the coordinated members of a finite set; similarly for $Verb$.

(*i*) Now consider the following English sentence which matches the pattern when $n = 1$:

```
The vaccine worked
```

Provide example sentences that extend this sentence for the case when $n = 2$ and $n = 3$. [2 marks]

(*ii*) Assuming that these constructions are part of the English language, would this mean that English is a Context-Free Language? Justify your answer. [3 marks]

(*iii*) Design a grammar in Chomsky Normal Form, G_2, which generates the finite matches of the pattern. [3 marks]

(*iv*) Specify a Push Down Automaton, M_2, that recognises $L(G_2)$. [3 marks]

5 Formal Models of Language

A linguist produces the grammar $G = (\mathcal{N}, \Sigma, S, \mathcal{P})$ where:

$$\begin{array}{rcl} \mathcal{N} & = & \{\text{S, X, Y, V, C}\} \\ \Sigma & = & \{a, contagious, highly, virus\} \\ S & = & \text{S} \\ \mathcal{P} & = & \{\text{S} \rightarrow \text{a X, X} \rightarrow \text{Y } virus \mid virus, \text{ Y} \rightarrow \text{V C} \mid \text{C}, \\ & & \text{V} \rightarrow highly \text{ V} \mid highly, \text{ C} \rightarrow contagious \text{ C} \mid contagious\} \end{array}$$

(a) Draw all the trees with 4 leaves that can be derived from this grammar. [2 marks]

(b) Based on corpus data the linguist assigns probabilities to each rule in his grammar. Describe how the probability of a string is calculated from the rule probabilities. [2 marks]

A mathematician prefers to generate the strings of a language inductively. She defines a homomorphism: $\{(a, a), (c, contagious), (h, highly), (v, virus)\}$. She defines $L \subset \Sigma^*$ where $\Sigma = \{a, c, h, v\}$ using the following axioms and rules:

$$\frac{}{av} \text{ (a1)}$$

$$\frac{u_1 v}{u_1 c v} \text{ (r1) where } u_1 \in \Sigma^*$$

$$\frac{a c u_1}{a h c u_1} \text{ (r2) where } u_1 \in \Sigma^*$$

$$\frac{u_1 h u_2}{u_1 h h u_2} \text{ (r3) where } u_1, u_2 \in \Sigma^*$$

(c) Let $L_i = \{u \in L \mid length(u) \leq i\}$. Find all members of L_4 [3 marks]

(d) Describe L as a regular expression and specify a Deterministic Finite Automaton, M, such that $L(M) = L$. [4 marks]

(e) Provide an expression for the conditional entropy of X for L_5, where X is a random variable over Σ. A numerical value is not required. [5 marks]

(f) Suggest some hypotheses about human language processing that we could test based on the models mentioned in this question. Provide reasons for your hypotheses. [4 marks]

6 Further Graphics

(*a*) An implicit function based on a kernel regressor can be defined as $f(\mathbf{x}) = \sum_i \alpha_i k_i(\mathbf{x})$, where α_i are scalar coefficients, $k_i(\mathbf{x}_i) = e^{-||\mathbf{x}-\mathbf{x}_i||^2/\sigma^2}$ are kernels, and $\mathbf{x}_i$ are sample points.

(*i*) Can $\alpha_i > 0$ define a valid implicit function, why/why not? [1 mark]

(*ii*) For a point $\mathbf{x}$ on the surface, compute the surface normal, simplify as much as possible. [2 marks]

(*iii*) Assume the points $\mathbf{x}_i$ are sampled from a plane passing through the origin. Will the regressor based implicit function preserve the plane normal? Show it mathematically. If not, explain a solution to better approximate the plane with the kernel regressor in 1-2 sentences. [3 marks]

(*b*) We have two surfaces A and B, both of which are represented continuously. We would like to compute if they intersect. How would you represent A and B (parametric and/or implicit) and why? Explain in 1-2 sentences. [2 marks]

(*c*) (*i*) Prove that a plane has zero mean curvature by utilizing $\nabla_{\mathcal{M}}\mathbf{p} = -2H\mathbf{n}$. [2 marks]

(*ii*) Given a triangular mesh with equilateral triangles and a vertex with non-negative discrete minimum and maximum curvatures, what is the maximum number of neighbours it can have? [3 marks]

(*d*) Assume we embed a number of bones inside the sphere of radius 1 and with center at the origin and start trying to deform the sphere by only having rotations at the bones.

(*i*) If we use linear blend skinning, show whether the transformed points can be off the sphere. [1 mark]

(*ii*) If we use linear quaternion blending with normalization, show whether the transformed points can be off the sphere. [2 marks]

(*iii*) If we use linear quaternion blending with normalization, will the sphere stay intact? Briefly explain. [2 marks]

(*iv*) If we use linear quaternion blending with normalization, but this time the sphere we are deforming is centered at $[1, 1, 1]^T$, show whether the transformed points can be off the sphere. [2 marks]

7 Further Graphics

(*a*) Recall that the (local) rendering equation is given as $L_o(\mathbf{x}, \vec{\omega}_o) = L_e(\mathbf{x}, \vec{\omega}_o) + \int_{H^2} f_r(\mathbf{x}, \vec{\omega}_i, \vec{\omega}_o) L_i(\mathbf{x}, \vec{\omega}_i) \cos\theta_i d\vec{\omega}_i$. Simplify this expression progressively as much as possible after each of the following assumptions. [5 marks]

(*i*) There is no light emitted from this surface point $\mathbf{x}$.

(*ii*) Diffuse reflection.

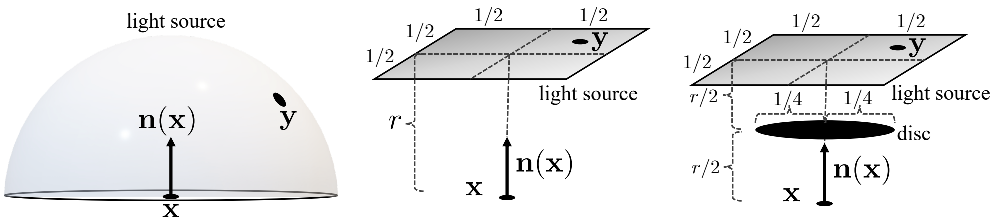
(*iii*) Reflection is constant at all surface points.

(*iv*) The incoming light is the same for all incident angles but there can be occluders.

(*v*) There is a single object and we ignore self-occlusions.

(*b*) Under certain assumptions, we can simplify the rendering equation to the following: $L_o(\mathbf{x}, \vec{\omega}_o) = f_r \int_{H^2} L_i(\vec{\omega}_i) \cos\theta_i d\vec{\omega}_i$. Assume we have a single known geometry for which we can compute surface normals $\mathbf{n}(\mathbf{x})$. Given three measurements from the surface $c_k = L_o(\mathbf{x}_k, \vec{\omega}_o)$ at known points $\mathbf{x}_k$, compute the reflected radiance $L_o(\mathbf{x}, \vec{\omega}_o)$ at an arbitrary point $\mathbf{x}$ on the surface. [*Hint:* If you encounter a linear system $\mathbf{Ax} = \mathbf{b}$, assume $\mathbf{A}^{-1}$ is known.] [4 marks]

(*c*) Recall that for direct illumination, we have the following form for the reflection equation: $L_r(\mathbf{x}, \mathbf{z}) = \int_A f_r(\mathbf{x}, \mathbf{y}, \mathbf{z}) L_i(\mathbf{x}, \mathbf{y}) G(\mathbf{x}, \mathbf{y}) dA(\mathbf{y})$. Starting from this form, provide answers to the following in terms of the provided parameters.



(*i*) Assuming a surface point $\mathbf{x}$ lit by a hemispherical light source with constant emitted light radiance L_i (left in the figure), derive the equation for the reflected light at $\mathbf{x}$ if the BRDF is a constant c. [*Hint:* $\int \cos\theta \sin\theta d\theta \equiv -0.5\cos^2\theta$] [4 marks]

(*ii*) Assuming a surface point $\mathbf{x}$ with a constant BRDF c lit by a planar 1×1 square light source distance r away (middle of the figure) where the emitted light radiance decays as $L_i(\mathbf{x}, \mathbf{y}) = 1/\cos^4\theta$ with θ the angle between the emitted light direction and plane normal, derive the equation for the reflected light at $\mathbf{x}$ (the surface normal $\mathbf{n}(\mathbf{x})$ is orthogonal to the planar light source). [4 marks]

(*iii*) Now assume we put an occluding opaque disc of radius 0.25 parallel to the light plane and centred at $\mathbf{x}$ between the light plane and $\mathbf{x}$ at a distance $r/2$ from $\mathbf{x}$ (right in the figure). Derive the equation for the reflected light at $\mathbf{x}$. [3 marks]

8 Further HCI

Predictive text entry systems are familiar on touch screens and mobile phones. This question asks you to consider how the same principles might be used in a programming editor for creating Java code.

(*a*) Explain using Bayes' Theorem how such an editor might update its expectation about which identifier will appear next, as the programmer is typing a line of code. Your answer should include a short extract of Java code preceding this line, in order to illustrate the context in which the expectation is being calculated. [5 marks]

(*b*) Describe how you could obtain an empirical measurement of the actual improvement in efficiency that results from using this predictive editor in practice. [5 marks]

(*c*) Consider the *four* possible combinations of a) small versus large variance, and b) small versus large effect size, in repeating this measurement. For *each* combination, explain the practical interpretation of that data for future development of the programming editor. [5 marks]

(*d*) If programmers have the option of whether or not to turn on this new function in the editor, describe some of the factors that might influence their decision-making process, with specific reference to the cognitive processes and sources of information that they would use. [5 marks]

9 Further HCI

(*a*) Augmented reality displays such as the Microsoft Hololens allow elements of the display to be registered against objects in the physical world. Briefly explain how such a display could be used for an Internet of Things application, setting different temperatures for multiple rooms in a home heating system. [1 mark]

(*b*) Explain how this compares to the use of visual metaphor on a conventional display for the same application, taking account of the correspondence between graphical resources and their representational purpose. [3 marks]

(*c*) Describe an alternative approach to visual representation that could be more efficient for setting desired temperature in a large number of rooms, for example in a multi-storey office building. Your description should mention three different categories of graphic resource, explaining the design principles for each of them in relation to this application. [8 marks]

(*d*) Choose three cognitive dimensions that are especially relevant to this problem, and explain how these dimensions would be different if the temperature settings were being adjusted using a speech interface such as Amazon Alexa or Google Assistant. [6 marks]

(*e*) How could an ethnographic approach be used to develop an alternative theory of correspondence relevant to this problem? [2 marks]

10 Prolog

A *Caesar Cipher* (or Shift Cipher) produces ciphertext from plaintext by replacing each letter with another that is found a fixed number of places down the alphabet. Users provide a key from 1 to 25 (inclusive) to determine the number of places to move. Our alphabet contains just the 26 lowercase letters and is circular: moving past `z` takes you back round to `a` again. For example under a key of 5 the letter `y` would be replaced by the letter `d`.

When answering this question ensure that each of your predicates has a comment giving a declarative reading of its behaviour and avoid unnecessary use of cut. Do not use any extra-logical predicates (such as `assertz`) or any library predicates.

(*a*) One way to represent the ordering of characters is with 26 facts indicating the next character. For example `next(a,b)` then `next(b,c)` through to `next(z,a)`.

Use `next` to implement a predicate `nextn(N,C1,C2)` which succeeds if the character `C2` appears `N` places after the character `C1`. You may assume that `N` is always a ground term. [3 marks]

(*b*) Another approach would be to use a list of characters to record the order of letters.

Provide an alternative implementation of `nextn` which makes use of the list representation `[a,b,c,...]`. Explain how you deal with the case of moving past the end of the alphabet.

You may assume the existence of two predicates: `scan(C,R,List)` which succeeds if `R` is the remainder of `List` that follows the letter `C`; and `charAt(N,C1,List)` which succeeds if `C1` is the character at position `N` in `List`. Position 0 is the first element of the list. `N` must be a ground term. [6 marks]

(*c*) Compare the merits of these two representations giving three relative benefits or drawbacks. [3 marks]

(*d*) Implement a predicate `caesar(K,P,C)` which succeeds if `C` is the ciphertext of the plaintext `P` under key `K`. Both ciphertext and plaintext are represented with a list of letters. You may assume that `P` and `K` are ground terms. [3 marks]

(*e*) The plaintext for a single ciphertext character has been discovered through a known-plaintext attack. Extend your `caesar` predicate to recover the key in this scenario and give an example invocation. [5 marks]

END OF PAPER