COMPUTER SCIENCE TRIPOS Part Ib

Tuesday 7 June 2005 1.30 to 4.30

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 Coversheets
Tags

1 Compiler Construction

(*a*) A (phrase-structured) grammar is often defined to be a 4-tuple (N, T, R, S) where R is a set of production rules. Explain what the other components of the 4-tuple are. Explain also the (most general form of) production rules, how these are conventionally restricted and why one might wish to restrict them. [6 marks]

(*b*) Give a grammar which is ambiguous. [2 marks]

(*c*) Give a grammar which is not a regular grammar but which generates a regular language containing an infinite number of strings. [2 marks]

(*d*) Is it possible to write a grammar which generates the strings $\{aa, aaa, aaaaa, \ldots, a^p, \ldots\}$ where p is prime? (A general argument for or against suffices.) [2 marks]

(*e*) It is desired to construct a simple "pocket-calculator" program using yacc and lex (or other similar automated tools of your choice) which can parse strings such as "`1+(10-5-3)*5+2=`" and print the result, 13 in this case. Outline the overall structure of your program components. Give full details of the input to yacc and lex (or equivalent). (Precise syntactic details are not important, but your answer should show an understanding of the principles involved.) [8 marks]

2 Economics and Law

(*a*) In auction theory, what is the difference between *strategic equivalence* and *revenue equivalence*? [8 marks]

(*b*) Give *three* examples of cases where revenue equivalence can fail, discussing in *each* case the implications for on-line auctions. [12 marks]

3 Data Structures and Algorithms

(a) A closed hash table is one in which the overflow chains of key–value pairs are held within the table itself. Carefully describe how the closed hash table mechanism works for both insertion and lookup. [6 marks]

(b) Assume that the initial probe is $p_0 =$ `Hash1(key)` `mod` B and the secondary probes are $p_i, i = 1 \ldots B-1$. Discuss the relative merits of the following schemes for choosing the secondary probes.

(i) $p_i = (p_0 + i)$ `mod` B

(ii) $p_i = (p_0 + 13 * i)$ `mod` B

(iii) $p_i = (p_0 + 13 * i + 17 * i * i)$ `mod` B

(iv) $p_i = (p_0 +$ `Hash2(key)` $* i + 17 * i * i)$ `mod` B

You may assume that all the arithmetic is unsigned. [8 marks]

(c) Carefully describe a mechanism for deleting key–value pairs from a closed hash table. [6 marks]

4 Artificial Intelligence I

A *perceptron* computes the function $h(\mathbf{x}) = \text{sgn}(\mathbf{w}^T\mathbf{x} + w_0)$ where $\text{sgn}(x) = +1$ if $x \geq 0$ and $\text{sgn}(x) = -1$ otherwise. The *primal perceptron algorithm* is as follows:

```
do
{
   for (each example in s)
   {
      if (y_i(w^T x_i + w_0) ≤ 0)
      {
         w = w + η y_i x_i
         w_0 = w_0 + η y_i R^2
      }
   }
}
while (mistakes are made in the for loop)
```

where η is a positive real, $R = \max_i ||\mathbf{x}_i||$ and $\mathbf{w}$ and w_0 are initialised to be the zero vector and 0 respectively, and where

$$\mathbf{s} = ((\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_m, y_m)) \text{ with } y_i \in \{+1, -1\}$$

is a training sequence.

(*a*) Derive the dual form of the perceptron algorithm and state the corresponding alternative representation for $\mathbf{w}$. [6 marks]

(*b*) Explain how the perceptron may be applied to problems that are not linearly separable by introducing *basis functions*. [4 marks]

(*c*) Give a definition of a *kernel*. [3 marks]

(*d*) Explain how the use of a suitable kernel in conjunction with the dual form of the perceptron algorithm can be advantageous compared with the direct use of basis functions and the primal perceptron algorithm. [7 marks]

5 Operating Systems II

(*a*) Scheduling disk requests is important to reduce the average service time. Describe briefly how the SSTF, SCAN and C-SCAN scheduling algorithms work. [3 marks]

(*b*) Recently there have been proposals that *2-D disk scheduling* algorithms should be used. What is the basic idea behind 2-D disk scheduling? [2 marks]

(*c*) You are asked to develop support for 2-D disk scheduling in a commodity operating system.

(*i*) What would be the major difficulty that you face? [2 marks]

(*ii*) Sketch a design for how you would go about overcoming this difficulty, and comment on how well you think the resulting system would work. [8 marks]

(*d*) Several modern file-systems and databases make use of a *journal* to aid in crash recovery.

(*i*) Briefly describe how journalling helps crash recovery. [2 marks]

(*ii*) A researcher suggests adding 128 MB of NVRAM to a disk drive and using this to store the journal. Discuss the advantages and disadvantages of this approach. [3 marks]

6 Continuous Mathematics

(a) Let $f(x)$ be a periodic function of period 2π. Give expressions for the Fourier coefficients a_r ($r = 0, 1, \ldots$) and b_r ($r = 1, 2, \ldots$) of $f(x)$ where

$$\frac{a_0}{2} + \sum_{r=1}^{\infty} (a_r \cos rx + b_r \sin rx)$$

is the Fourier series representation of $f(x)$. [2 marks]

(b) Show that the Fourier series in part (a) can also be written as a complex Fourier series

$$\sum_{r=-\infty}^{r=\infty} c_r e^{irx}$$

by deriving expressions for the complex Fourier coefficients c_r ($r = 0, \pm 1, \pm 2, \ldots$) in terms of a_r and b_r. [3 marks]

(c) Use your expressions for a_r and b_r in part (a) and for c_r in part (b) to show that

$$c_r = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-irx} dx \qquad (r = 0, \pm 1, \pm 2, \ldots)\,.$$

[3 marks]

(d) Show that the complex Fourier coefficients of $f(x - \alpha)$ (where α is a constant) are given by $e^{-ir\alpha} c_r$ ($r = 0, \pm 1, \pm 2, \ldots$). [6 marks]

(e) Suppose that $g(x)$ is another periodic function of period 2π with complex Fourier coefficients d_r ($r = 0, \pm 1, \pm 2, \ldots$) and define $h(x)$ by

$$h(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x - y) g(y) dy\,.$$

Show that $h(x)$ is a periodic function of period 2π and that its complex Fourier coefficients are given by $h_r = c_r d_r$ ($r = 0, \pm 1, \pm 2, \ldots$). [6 marks]

[*You may assume that the periodic functions in this question satisfy the Dirichlet conditions. Euler's equation may be used without proof but should be stated precisely.*]

7 Numerical Analysis I

(*a*) Define *absolute error* and *relative error*. How are these related? Explain briefly the term *loss of significance*. [3 marks]

(*b*) An algorithm is required for solution of $ax^2 + bx + c = 0$ where $b > 0$. Describe how loss of significance can occur in the formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

and derive an alternative formula for such a case. Illustrate your answer by applying it to the case $a = 30$, $b = 3000$, $c = 1$ on a decimal machine with only 5 significant digits available. [You should assume for the purposes of calculation that $\sqrt{b^2 - 4ac}$ evaluates to $b - (2ac/b)$, correctly rounded.] [10 marks]

(*c*) The series

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots$$

is to be summed by taking terms in order, left to right, using only p decimal digits of precision until additional terms are negligible. If $x = 6$ find the largest term of the series and hence, assuming $\cos 6 \simeq 1$, estimate roughly how many decimal digits of accuracy will be lost in the process. [7 marks]

8 Concurrent Systems and Applications

(*a*) Java includes support for converting between objects and a stream of bytes.

(*i*) Describe the Java serialization mechanism. [4 marks]

(*ii*) What circumstances can cause exceptions to be thrown during serialization? [2 marks]

(*iii*) What is the effect of the `transient` modifier? To which components of a class is it legal to apply the `transient` modifier? [3 marks]

(*iv*) Why do some programmers prefer to use `Externalizable` instead of `Serializable`? [1 mark]

(*b*) Java includes support for converting a stream of bytes into a class definition. Describe the facilities that provide this functionality and ensure that the type-safety of the Java Virtual Machine cannot be compromised. [4 marks]

(*c*) What are *native methods* and why might they be useful? [2 marks]

(*d*) What are *reference objects*? How does a *ReferenceQueue* provide a means to release resources when they are no longer needed? Explain how reference objects can be used to construct data structures that permit the garbage collector to release memory when the Java Virtual Machine might otherwise run out of memory. [4 marks]

9 Computation Theory

(a) What does it mean for a subset S of the set $\mathbb{N}$ of natural numbers to be register machine *decidable*? [3 marks]

(b) For each $e \in \mathbb{N}$, let $\varphi_e \in \mathit{Pfn}(\mathbb{N}, \mathbb{N})$ denote the partial function computed by the register machine program with index e. Let $e_0 \in \mathbb{N}$ be an index for the totally undefined partial function (so that $\varphi_{e_0}(x)\uparrow$, for all $x \in \mathbb{N}$).

Suppose that a total function $f \in \mathit{Fun}(\mathbb{N}, \mathbb{N})$ is *extensional*, in the sense that for all $e, e' \in \mathbb{N}$, $f(e) = f(e')$ if φ_e and $\varphi_{e'}$ are equal partial functions. Suppose also that the set $S_f = \{x \in \mathbb{N} \mid f(x) = f(e_0)\}$ is not the whole of $\mathbb{N}$, so that for some $e_1 \in \mathbb{N}$, $f(e_1) \neq f(e_0)$.

(i) If membership of S_f were decided by a register machine M, show informally how to construct from M a register machine M' that, started with $R1 = e$ and $R2 = n$ (any $e, n \in \mathbb{N}$) always halts, with $R0 = 0$ if $\varphi_e(n)\downarrow$ and with $R0 = 1$ if $\varphi_e(n)\uparrow$. Make clear in your argument where you use the fact that f is extensional.

[Hint: For each $e, n \in \mathbb{N}$ consider the index $i(e, n) \in \mathbb{N}$ of the register machine that inputs x, computes $\varphi_e(n)$ and if that computation halts, then computes $\varphi_{e_1}(x)$.] [14 marks]

(ii) Deduce that if f is extensional, then S_f is either the whole of $\mathbb{N}$, or not decidable. [3 marks]

END OF PAPER