# COMPUTER SCIENCE TRIPOS Part II (General)
# DIPLOMA IN COMPUTER SCIENCE

---

Wednesday 6 June 2007 1.30 to 4.30

---

PAPER 12 (PAPER 3 OF DIPLOMA IN COMPUTER SCIENCE)

*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
*None*

## 1 Natural Language Processing

($a$) In German, the third person singular present inflection of weak verbs is generally formed by adding the 't' to the stem. Exceptions to this rule include verbs with stems that end in 't' or 'd' which are formed by adding 'et' instead of 't'. The following table gives some examples (^ is used as the affix marker in the underlying form).

| stem | surface | underlying |
|---|---|---|
| kauf | kauft | kauf^t |
| arbeit | arbeitet | arbeit^t |

Draw a finite state transducer (FST) that relates surface and underlying forms according to this pattern. (Only the inflected forms should be accepted by the transducer since the stems by themselves do not correspond to words.) Explain the notation that you use and outline how the FST could be used in morphological analysis and generation. [14 marks]

($b$) The past participle of these verbs is the same as the third person singular present, but with 'ge' before the stem. Assume that the underlying form of the past participle is treated as having the artificial suffix '^P', as indicated below.

| stem | surface | underlying |
|---|---|---|
| kauf | gekauft | kauf^P |
| arbeit | gearbeitet | arbeit^P |

Modify the FST that you gave in answer to part ($a$) above to allow for the past participle as well as the third person singular present. What does this example illustrate about limitations of the FST approach? [6 marks]

## 2 Computer Design

(*a*) What is the difference between the *latency* and the *bandwidth* of a communication link? [2 marks]

(*b*) Why are control-flow processors sensitive to memory latency even if the memory bandwidth exceeds what the processor could ever require? [4 marks]

(*c*) How is average memory latency typically reduced? [5 marks]

(*d*) What is the difference between serial and parallel communication and what is the receiver required to do to recover the data in each case? [5 marks]

(*e*) The PCI communication standard for add-on cards for PCs has moved to PCI express (PCIe). PCI uses a parallel bus to communicate to several cards whereas PCIe uses sets of point-to-point serial links for communication. Why are multiple serial links rather than parallel buses preferred for high bandwidth communication? [4 marks]

## 3 Digital Communication I

Using *four* examples, explain how multiple higher-layer channels can be multiplexed onto a lower-layer channel. In each example consider

(*i*) how the individual higher-layer channels can be recognised;

(*ii*) what the mechanism is for allocation of lower-layer channel resources to the higher-layer channels; and

(*iii*) the characteristics of the higher-layer channels.

[$4 \times 5$ marks]

## 4 Distributed Systems

(*a*) Define the characteristics of middleware based on the following interaction paradigms:

(*i*) request-reply;

(*ii*) one-to-one message passing.

Discuss any limitations imposed by the paradigms and mention the properties of the application domains for which each is particularly suited. [9 marks]

(*b*) Web Services standards have been specified relatively recently. What does the Web Services paradigm offer? [3 marks]

(*c*) (*i*) Define the publish/subscribe communication paradigm. [2 marks]

(*ii*) What does this paradigm provide that is lacking in the three described above in parts (*a*) and (*b*)? [3 marks]

(*iii*) What are the shortcomings of publish/subscribe? [3 marks]

## 5 Computer Graphics and Image Processing

(*a*) Describe, in detail, an algorithm that will draw a one-pixel wide outline of a circle of integer radius centred on the origin. [10 marks]

(*b*) Describe the modifications required to your algorithm in part (*a*) to make it draw a filled circle. [3 marks]

(*c*) Describe the modifications required to your algorithm in part (*a*) to make it draw the outline of a circle centred at arbitrary integer coordinates. [2 marks]

(*d*) Describe the modifications required to your algorithm in part (*c*) to make it draw the outline of a circle centred at arbitrary non-integer coordinates and of non-integer radius. [5 marks]

## 6 Compiler Construction

(*a*) Describe how a stack is used to implement procedures and functions. [6 marks]

(*b*) Suppose a language allows the creation of pointers. How does this complicate the use of stacks as described in part (*a*)? [2 marks]

(*c*) How does the Java language deal with the problem described in part (*b*)? [2 marks]

(*d*) Consider the following ML-like program containing the function `g` that returns a function as a result.

```
let a = 17 in
let g b = (let h c = a + b + c in h) in
let f1 = g 21 in
let f2 = g 33 in
let v = f1(3) + f2(57) in
...
...
```

Explain carefully how such a program can be compiled. In particular, pay special attention to how the code for the body of the function `h` can access the values of `a`, `b`, and `c`. [10 marks]

## 7 Concepts in Programming Languages

(*a*) Give an overview of the FORTRAN execution model (or abstract machine) and comment on its merits and drawbacks from the viewpoints of programming, compilation, execution, etc. [5 marks]

(*b*) What is *garbage collection* in the context of programming languages? Comment on its merits/drawbacks and explain the contexts in which it arises, giving examples. [5 marks]

(*c*) Recall that Algol has a primitive static type system. In particular, in Algol 60, the type of a procedure parameter does not include the types of its parameters. Thus, for instance, in the Algol 60 code

```
procedure P( procedure F )
    begin F(true) end ;
```

the types of the parameters of the procedure parameter `F` are not specified.

Explain why this piece of code is statically type correct. Explain also why a call `P(Q)` may produce a run-time type error, and exemplify your answer by exhibiting a declaration for `Q` with this effect.

Why does this problem not arise in SML? [5 marks]

(*d*) Give an implementation of the abstract data type of stacks using the SML module system. [5 marks]

## 8 Databases

(*a*) The Entity/Relationship model is based around the concepts of *entity*, *attribute*, and *relationship*. Describe how these can be represented in the relational model. [6 marks]

(*b*) Data normalisation is often an important component in database design. Discuss why this is so, and give examples of situations where normalisation is *not* important. [6 marks]

(*c*) Let $A$ and $B$ be disjoint non-empty sets of attributes. Let $R$ be a relation over attributes $A \cup B$ and let $S$ be a relation over attributes $B$.

Suppose that we want to introduce a new relational operation called *division*, denoted $R \div S$, that will return a relation over attributes $A$. The relation $R \div S$ is made up of all tuples $t$ such that for all $s \in S$ we have $ts \in R$ ($ts$ is the concatenation of $t$ and $s$).

Note that in the special case that $R = T \times S$ for some relation $T$, then $(R \div S) = T$ and $(R \div T) = S$.

In other words, $\div$ can be treated as an inverse to the Cartesian product.

Can we define $R \div S$ in the relational algebra? Prove that your answer is correct. [8 marks]

## 9 Quantum Computing

(*a*) Give a schematic circuit diagram for Grover's algorithm. [4 marks]

(*b*) Suppose we apply Grover's algorithm to a four-qubit register in which exactly one of the states is marked.

(*i*) Writing $O$ for the oracle matrix, write out the Grover iterate in matrix form. [4 marks]

(*ii*) What are the probabilities of measuring the marked state after applying the Grover iterate *once*, *twice* and *three* times? [3 marks each]

(*iii*) If you continue to increase the number of times the Grover iterate is applied, will the probability continue to increase? Justify your answer. [3 marks]

## 10 Bioinformatics

(*a*) Describe with *one* example the difference between Hamming and Edit distances. [2 marks]

(*b*) Discuss the Smith–Waterman algorithm. What is the complexity and the relationship with the problem of finding the longest common subsequences? [5 marks]

(*c*) Describe the Banded algorithm for local alignment and its complexity. [5 marks]

(*d*) Describe the four Russian speedup algorithm. [8 marks]

## 11 Software Engineering and Design

What is the relationship between requirements and testing in the following contexts?

(*a*) the Unified Modeling Language

(*b*) user-centred design

(*c*) agile development methods such as eXtreme Programming

(*d*) module design

(*e*) ISO 9000 quality certification

(*f*) the waterfall development model

[up to 4 marks each, maximum 20]

## 12 Complexity Theory

($a$) Give a precise definition of polynomial-time reductions. [2 marks]

($b$) Give a precise definition of NP-completeness. [3 marks]

($c$) Let **Subset Sum** denote the following decision problem:

Given a set of positive integers $S = \{v_1, \ldots, v_n\}$ and a number $t$, determine whether there is a subset of $S$ that sums to exactly $t$.

($i$) Explain why **Subset Sum** is in NP. [3 marks]

($ii$) Describe a polynomial-time reduction from the problem of 3-dimensional matching to **Subset Sum**. [9 marks]

($iii$) Explain why parts ($i$) and ($ii$) above imply that **Subset Sum** is NP-complete. [3 marks]

**END OF PAPER**