University of Toronto
csc228S - File Structures and Data Management, Spring 2000

Midterm Test

Friday February 25, 2000

Duration: 50 minutes

Aids allowed: None

Family Name: __________________ Given names: __________________

Student #: __________________ Tutor: __________________

- There are 6 pages, including this one. The test is out of 35 marks and the value of each question is provided; please use this information to manage your time effectively.
- For questions that involve writing code, comments are not necessary. If you need to call a standard method but can't remember the correct order of arguments, just indicate the meaning of each argument.

Part A:	______	/ 5
Part B:	______	/ 6
Part C:	______	/ 4
Part D:	______	/ 6
Part E:	______	/ 5
Part F:	______	/ 4
Total	______	/ 30

Part A [5 marks in total; **no marks** will be given without the "Tree / Explanation"]

Here is a table of characters and their corresponding frequency in a file we would like to compress.

character	frequency
A	60%
B	30%
C	5%
D	2%
E	3%

Below are five different possible encodings. For each one, circle whether or not it could be a Huffman encoding. If you circle "yes", draw the corresponding tree. If you circle "no", explain why it is not possible for the encoding to be a Huffman encoding.

character	code
A	1
B	11
C	110
D	1110
E	1111

YES NO

Tree / Explanation:

character	code
A	0
B	11
C	101
D	1001
E	1000

YES NO

Tree / Explanation:

character	code
A	110
B	10
C	1110
D	0
E	1111

YES NO

Tree / Explanation:

character	code
A	00
B	01
C	10
D	111
E	110

YES NO

Tree / Explanation:

character	code
A	1
B	01
C	001
D	0001
E	0000

YES NO

Tree / Explanation:

Part B [6 marks in total]

Consider 3 files of records with very different layouts.

File A Contains records about students, with the following fixed length fields, in this order:

StudentName	30 characters, left justified, padded with blanks
ID	9 digits (represented as text)
GPA	4 characters xx.x where x is a digit 0..9

There are no delimiters between fields or between records.

File B Contains records about companies, with the following fields, in this order:

CompanyName	variable length, character data
Address	variable length, character data
PostalCode	6 characters
ContactName	variable length, character data

There is a delimiter after each field and no additional delimiter between records.

File C Contains records about images, where each record has the following fields, in this order:

ImageNameLength	5 bits	indicates the length of the ImageName field
ImageName	variable	character data
Xoffset	4 bits	(value 0..15)
Yoffset	4 bits	(value 0..15)
Scale	4 byte	floating point value, in binary format

There are no delimiters between fields or between records.

1. In File B, the actual character for the delimiter is not stated. The designer could choose to use a readable or non-readable character for this. State one advantage for each alternative.
   ADVANTAGE OF NON-READABLE DELIMITER:

   ADVANTAGE OF READABLE DELIMITER:

2. Which field in any of the files might contribute to internal fragmentation?

3. In File B, which delimiters could be removed without making other changes to the file format? Circle the one correct answer.
   (a) those between CompanyName and Address
   (b) those between Address and PostalCode
   (c) those between PostalCode and ContactName
   (d) those between ContactName and CompanyName
   (e) all the delimiters are needed

4. State the limit on the number of characters in the ImageName field in File C. ________

5. Which file could have records referenced by RRN? Circle one. **A** **B** **C**

Part C [4 marks in total]

Suppose an operating system keeps track of the blocks of a file using a multi-level index with 3 levels. Each block of the index holds 2^{10} bytes and a pointer to a file block takes 4 bytes. What is the size of the biggest file possible in this file system?

ANSWER: ____________ blocks

ROUGH WORK:

Part D [6 marks in total]

Suppose we have a file of integers, in binary format. Write a fragment of C++ code that goes to the integer that is 6 integers away from the beginning of the file, reads that integer, and then prints it to the standard output. The code to open the file is already written.

```
fstream file;
file.open("fileOfInts", ios::in)
```

Part E [5 marks in total]

Below is a small program, and the statement used to compile it. It won't compile. Make any changes necessary below so that it will compile and run.

File Node.h

```
template <class X>
class Node {
public:
    Node<X> () { data=0; next=0; };

    // Set my data field to "value".
    void set(X * value);

    // Return my data field.
    X * get();

private:
    X * data;
    Node<X> * next;
};
```

File Node.cpp

```
#include "Node.h"

template<class X> void Node<X>::set (X * value){
    data = value;
};

template<class X> X * Node<X>::get (){
    return data;
};
```

File Client.h

```
class Client {
public:
    int accountNumber;
    float balance;
};
```

File driver.cpp

```
#include <fstream.h>
#include "Client.h"
#include "Node.h"

int main(void) {
    Client c1;
    c1.accountNumber = 91524;
    c1.balance = 0.0;
    Node<Client> n;
    n.set(c1);
    cout << n.get();
    return 0;
}
```

COMPILING COMMAND: `g++ driver.cpp`

Part F [4 marks in total]

Suppose we have a B-tree of order 5 (i.e. $M = 5$), and that the tree has nodes on three levels — a root node and two levels below it.

(a) What is the minimum number of leaf nodes that this tree might have? ______________

(b) What is the *maximum* number of leaf nodes that this tree might have? ______________

(c) What is the minimum number of keys that this tree might have at the leaf level? ______________

(d) What is the *maximum* number of keys that this tree might have at the leaf level? ______________

End of Examination