UNIVERSITY OF TORONTO
Faculty of Arts and Science

April/May Examinations 1996

CSC228S

Duration — 3 hours

Aids allowed: One 8.5 by 11 inch aid sheet, as described in class

- **Make sure your examination booklet has 12 pages (including this one).**
- Write your answers in the spaces provided. Do not feel that you must use all of the space provided. If you run out of space on any question, use the back of a page, and draw an arrow to point this out.
- You will be rewarded for concise, well-thought-out answers, rather than long rambling ones.
- Write legibly. Unreadable answers will be given 0.

Family Name: ______________________ Given Names: ______________________

Student #: ______________________ Tutor: ______________________

1. ________ / 8
2. ________ / 10
3. ________ / 19
4. ________ / 10
5. ________ / 13
6. ________ / 8
7. ________ / 10
8. ________ / 17

Total ________ / 95

Question 1

[8 marks in total]

Note: You may find the conversion tables at the bottom of this page helpful in answering this question.

To a UNIX system, text files and binary files look the same; they are both just sequences of bytes. The difference is in how the bytes are used to stand for things like words and numbers. For this question, assume that integers are represented by four bytes in memory. Specify any other assumptions that you need to make.

Assume that the following C code has been executed:

```
int num;
num = 41;
FILE *fp;
fopen (fp, "out", w);
```

a. Write a single statement in C that will write `num` to file `out` in binary.

b. Show below the sequence of bytes that would be written to the file as a result of your C statement from part *a*.

c. Write a single statement in C that will write `num` to file `out` as text.

d. Show below the sequence of bytes that would be written to the file as a result of your C statement from part *c*.

character	ASCII value
1	49
4	52
\ n	10

decimal number	binary equivalent
1	00000001
4	00000100
10	00001010
41	00101001
49	00110001
52	00110100

Question 2

[10 marks]

Write a C function that takes an array of integers, an integer n, and a `FILE *` that points to an already open file, and writes each of the first n integers from the array out to the file in binary. After each integer, your function should write a file position indicator that points to the previous integer, or `NULL`, in the case of the very first integer. Your function should return a pointer to the last integer in the file, which is at that point, the “front” of a linked list of integers.

Assume that $n \geq 1$, and that nothing goes wrong.

Question 3

[19 marks in total]

a. [3 marks]
Draw a B-tree of order 5, and with two levels (*i.e.*, a root, and one level below it). Let each of the records stored within the nodes consist only of an integer key. Use the smallest allowable number of nodes, and the smallest allowable number of records per node.

b. [2 marks]
Give two reasons why, in a B-tree, there is a minimum number of children per node.

c. [2 marks]
Why do B-trees grow upwards from the root, rather than downwards from the leaves?

d. [2 marks]
Consider a B-tree with branching factor 7. What is the minimum number of pointers required in every internal node in such a tree?

What is the maximum?

e. [3 marks]
Imagine that you are performing an insertion in a 7-way B-Tree. The new value is to go in a leaf node, but the node is already full. In such cases, the leaf node is "split." Describe exactly what this involves, in the simplest case. Include a diagram.

f. [5 marks]
Prove that, in an 7-way B-tree, after the leaf node is split, the resulting new leaves contain at least the minimum required number of pointers. This does not require any fancy proof techniques.

g. [2 marks]
What repercussions may occur beyond the leaf level? Explain.

Question 4

[10 marks in total]

a. [4 marks]

Recall double hashing, one of the collision resolution schemes that we discussed. What problem of linear-hashing is double hashing designed to solve, and how does it solve that problem?

b. [3 marks]

The operating system controls one thing that influences the choice of an appropriate bucket size when hashing. Explain what it is, and its influence.

c. [3 marks]

Consider the following two hash functions for a 12-digit key:

- hash function # 1: take the last 4 digits of the key.
- hash function # 2: take the sum of each group of 4 digits, and mod by 10000 to get a value between 0 and 9999.

Without knowing anything about the keys other than that they are 12-digit integers, which hash function would you prefer? Why?

Question 5

[13 marks in total]

Recall linear hashing, one of the extendible hashing schemes. When the performance of a linear hashing scheme degrades, a bucket is “split”, starting from bucket zero.

a. [3 marks]
Is Linear Hashing compatible with collision resolution by open addressing? Explain.

b. [4 marks]
Write a hash function that is suitable for use with linear hashing. Assume that, before any splitting, the hash function should return the key mod 13.

Your function will need more parameters than just the hash key; add them as needed, but clearly explain what each parameter is for. Pseudo-code is fine; you do not have to answer this question in C.

c. [6 marks]
Now write a function for splitting the next bucket. Assume collision resolution has been handled by chaining, and that you already have functions for inserting and deleting elements from given buckets.

Again, include and explain any parameters that you need. Pseudo-code is fine.

Question 6

[8 marks in total]

a. [1 mark]
What is the benefit of using table-assisted hashing? Be specific.

b. [4 marks]
Describe two extra costs associated with table-assisted hashing.

c. [3 marks]
Is table-assisted hashing designed to go with collision resolution by chaining, or by open addressing. Explain.

Question 7

[10 marks in total]

a. [3 marks]
Modify the diagram below to create the structure of a threaded file, indexed according to campus.

NAME	DEPARTMENT	CAMPUS
Liszt	Computer Science	Downtown
Kodaly	Physics	Downtown
Byrd	Music	Erindale
Bach	German	Downtown
Palestrina	Computer Science	Scarborough
Victoria	Philosophy	Scarborough
Josquin	Philosophy	Scarborough
Orff	Computer Science	Downtown
Haydn	Computer Science	Downtown
Purcell	German	Scarborough
Handel	Physics	Erindale

b. [2 marks]
What is the difference between a threaded file and a multilist?

c. [3 marks]
What advantage does a multilist have over a threaded file? Explain how it gets this advantage.

d. [2 marks]
What cost does a multilist have over a threaded file?

Question 8

[17 marks in total]

A video rental store has a relational database to keep track of their customers, videotapes, etc. Their database includes the files below. The RENTALS file keeps a history all past rentals, up to the current date.

file VIDEOS:

IDNumber	Title	Category
8818398	Sleepless in Seattle	Romance
1234562	Rambo	Action
2222222	Friday the 13th	Horror

... *etc.*

file CUSTOMERS:

MembershipNumber	Name	PhoneNumber	CreditCardNumber
12345	Diane Horton	555-1212	5191 8244 4568 9999
15111	Jane Davies	787-3724	5191 1111 2222 3333
99999	Lydia Veen	788-0924	5224 9813 3543 0153

... *etc.*

file RENTALS:

IDNumber	MembershipNumber	Year
8818398	99999	1996
8818398	12345	1995
2222222	15111	1995

... *etc.*

a. Write a query to list the video ID numbers of all movies that were rented in 1995.

b. Write a query to list the titles of all movies that were rented in 1995.

c. You want to contact customers who haven't been active lately. Write a query to make a mailing list of everyone who has not rented a movie this year. The query should produce their names and addresses.

d. You want to sell a mailing list to a company that markets Kleenex, a brand of facial tissue. Write a query to list the names and addresses of everyone who has rented a romance.

e. Describe two facts about the domain that are not *explicitly* represented by the database, but are implied, and underlie how we respond to queries.

END OF EXAM; Total pages = (12)