University of Ottawa
CSI 2131 – Midterm Test
Instructor: Lucia Moura

March 3rd, 2001
1:00 pm
Duration: 2 hs

Closed book, no calculators
Version I

Last name: ______________________

First name: ______________________

Student number: ______________________

Circle your section:

Section A (Mon and Thur) Section B (Tues and Fri)

There are 15 questions and 15 marks total.

This exam paper should have 17 pages,
including this cover page.

1 – File Processing in C++	/ 1
2 – Records and Fields	/ 1
3 – Disks	/ 1
4 – Tape	/ 1
5 – CD-ROM	/ 1
6 – Basic I/O Software and Hardware	/ 1
7 – Buffering Strategies	/ 1
8 – Lempel-Ziv Compression	/ 1
9 – Huffman Compression	/ 1
10 – Data Compression	/ 1
11 – Reclaiming Space	/ 1
12 – Fragmentation	/ 1
13 – Sorting and Searching	/ 1
14 – Indexing	/ 1
15 – Co-sequential Processing	/ 1
Total	/ 15

1 File Processing in C++ — 1 point

Consider the following part of a C++ program:

```
#include <fstream.h>
int main() {
  char ch = 'x';
  fstream f;
  f.open("myfile.txt",ios::in);
  .
  .
  .
}
```

Which of the following file processing operations is GUARANTEED TO RESULT IN ERROR if executed later on in this program?

A. `f << c;`

B. `f >> c;`

C. `f.seekg(0,ios::beg);`

D. `f.seekg(10;ios::cur);`

E. `f.close();`

2 Records and Fields — 1 point

Select the statement that is **FALSE**:

A. A primary key is a key that is used to identify uniquely a record.

B. A key may be the combination of one or more fields in a record.

C. Beginning a field with a length indicator or separating fields with delimiters are methods used in variable-length fields.

D. Fixed-length records may contain variable-length fields.

E. If a file contains variable-length records, the RRN (relative record number) can be used to calculate the byte offset of a record.

... continued

3 Disks — 1 point

You are given a 16,000 50-byte record file to place on secondary memory. Three disks are available that can store 25,600 bytes per track and 4 tracks per cylinder, but that use different organizations:

- Disk 1:
  - organized by sector
  - 128 bytes per sector
  - no record is allowed to span two sectors (though several records can be stored per sector)
- Disk 2:
  - organized by count-data block
  - blocking factor = 50
  - extra amount of space (used by count-subblock and interblock gaps) = 60 bytes
- Disk 3:
  - organized by count-data block
  - blocking factor = 10
  - extra amount of space (used by count-subblock and interblock gaps) = 300 bytes

Which of the following statements is **CORRECT**?

A. Disk 1 requires more cylinders to store the file than Disk 2 and 3.

B. Disk 1 requires fewer cylinders to store the file than Disk 2 but more than Disk 3.

C. Disk 1 requires more cylinders to store the file than Disk 2 but fewer than Disk 3.

D. All three Disks require as many cylinders to store the file.

E. Disk 1 requires fewer cylinders to store the file than either Disk 2 or 3.

4 Tape — 1 point

Consider the following tape description:

Density = 3,000 bpi

Interblock Gap = 0.2 inch

Speed = 200 ips

Suppose we are planning to store a 1,000 30-byte record file on this tape using a blocking factor of 10. What is the tape's **effective transmission rate**?

A. 1000 bpi

B. 200 Kilobytes/sec

C. 1 Megabyte/sec

D. 3000 bpi

E. 30 inches

5 CD-ROM — 1 point

Select the answer that describes a **STRENGTH** attributed to CD-ROMs:

A. Fast seek performance.

B. Fast transfer rate.

C. Large storage capacity.

D. CLV (constant linear velocity) organization.

E. Read/Write capability.

... continued

6 Basic I/O Software and Hardware — 1 point

Select the statement that is **FALSE**:

A. The I/O processor runs independently of the CPU, allowing for I/O processes and internal computing to overlap.

B. The disk controller is responsible for controlling the operation of the disk drive.

C. The following are tasks attributed to the operating system/file manager: keeping a file allocation table, managing the use of buffers, deciding which cluster on disk will be allocated to a file.

D. If at a given time the disk drive is available for writing, the system I/O buffer will not be used, since bytes can be written directly to the disk.

E. A cluster consists of one or more contiguous sectors. The use of large clusters can improve sequential access times, since longer spans of data can be read without seeking.

7 Buffering Strategies — 1 point

Which of the following statements is **CORRECT**:

A. There exists many different types of buffers including system buffers which handle I/O and program buffers which handle a program's data.

B. When using a Buffer Pooling strategy, it is advisable to use the MOST RECENTLY USED strategy to select a buffer for usage from the buffer pool. In this way, we can be sure that the buffers that have been used most recently do not remain in memory for too long.

C. DOUBLE BUFFERING, i.e. using 2 buffers, will slow the I/O operation down when a program needs to both read from the disk and write to the disk during its execution.

D. Buffering will make I/O operations more efficient only when I/O-CPU overlapping is NOT permitted.

E. When a program crashes, the content of the buffers that the program was using before the crash is always saved onto secondary storage.

... continued

8 Lempel-Ziv Compression — 1 point

Consider the following string

```
ababcaabcabc
```

If we assume that 8 bits are normally needed to encode each character, then how many bits does Lempel-Ziv encoding allow us to save with respect to the original message?

A. 0 bits

B. between 1 and 20 bits (included)

C. between 21 and 40 bits (included)

D. between 41 and 60 bits (included)

E. more than 60 bits

... continued

9 Huffman Compression — 1 point

Consider a file containing the following letters with their frequencies:

letter:	A	B	C	D	E
frequency:	40	20	15	14	11

Choose the only VALID HUFFMAN CODE for this file:

A.

letter:	A	B	C	D	E
CODE:	0	10	110	1111	1110

B.

letter:	A	B	C	D	E
CODE:	0	111	110	101	100

C.

letter:	A	B	C	D	E
CODE:	11	10	01	001	000

D.

letter:	A	B	C	D	E
CODE:	0	11	10	01	00

E.

letter:	A	B	C	D	E
CODE:	0	111	101	110	100

... continued

10 Data Compression — 1 point

Which of the following statements is **CORRECT**:

A. Compression is seldom worthwhile since the time it takes to run a compression program does not justify the amount of space saved by the compression.

B. One must consider the usual time/space trade-off when using data compression: while data compression allows the use of less disk storage, it increases the time necessary to transmit the file over a data channel such as the internet.

C. RUN-LENGTH ENCODING, the compression method that consists of summarizing repeated contiguous values, is more effective for text compression than for image compression.

D. Lempel-Ziv and Huffman Encoding are space efficient compression techniques routinely used in UNIX systems.

E. Huffman and Lempel-Ziv encoding are both able to reach the same expected number of bits needed to encode each symbol, but Lempel-Ziv encoding also requires knowledge of the symbols' probabilities.

... continued

11 Reclaiming Space — 1 point

Consider the following file containing variable-length records and fields:

```
26Smith|James|Dec|17|1973|
38Dupont-Sur-Seine|Michel|Feb|22|1970|
21Elk|Ann|Mar|6|1976|
31Shelling|Anthony|May|29|1969|
```

The record-length is indicated by a length indicator at the beginning of each record. Fields are separated by delimiter: '|'.

The file contains 4 records: R1 of size 26, R2 of size 38, R3 of size 21 and R4 of size 31.

(Note: in the actual file, no carriage-returns were inserted after each record. They were inserted here, only for the sake of readability)

We use a list of available records (AVAIL LIST) for reclaiming space from deleted records. Assume that a WORST-FIT strategy is used in order to reduce external fragmentation. Assume that the AVAIL LIST is initially empty.

Later, the above file is modified in the following order:

- record R3 is deleted,
- record R5 of size 25 is added,
- record R4 is deleted,
- record R6 of size 17 is added.

What is the content of the AVAIL LIST (in terms of the size of the blocks included in it) after these modifications have been performed?

A. 31, 21

B. 26, 38, 25, 17

C. 31

D. 21, 14

E. 26, 38, 17, 31, 25

... continued

12 Fragmentation — 1 point

Which of the following phrases is **WRONG**?

A. Wasted space within a record is called INTERNAL FRAGMENTATION.

B. Record deletion in variable-length record files may cause EXTERNAL FRAGMENTATION.

C. STORAGE COMPACTION can be used to combat external fragmentation.

D. FRAGMENTATION can be dealt with dynamically by reclaiming deleted space when records are added.

E. EXTERNAL FRAGMENTATION is usually WORSE when deleted records are of fixed-length as opposed to variable-length.

... continued

13 Sorting and Searching — 1 point

Choose the **CORRECT** answer:

A. One advantage of KEYSORTING is that it requires less memory than a regular internal sorting of a file.

B. INTERNAL SORTING of a file refers to any sorting method that takes place in main memory, including the situation in which the file does not completely fit into main memory.

C. SEQUENTIAL SEARCH on a file takes less time than BINARY SEARCH on the file, since reading a file sequentially is faster than seeking through the file.

D. A BINARY SEARCH on a file of n records (for arbitrary n) takes in the worst case n disk accesses.

E. INTERNAL SORTING of a file of n records (for arbitrary n) requires at least $n \times \log n$ accesses to secondary storage.

... continued

14 Indexing — 1 point

Consider a file of "recordings" similar to the one seen in class. The following is the **composer index**, which is a **secondary index** using **inverted lists**:

Secondary Key Index:

0	BEETHOVEN	2
1	DVORAK	3
2	PROKOFIEV	5

List of Primary Key References:

0	DG139201	4
1	LON2312	-1
2	ANG3795	0
3	COL31809	-1
4	RCA2626	-1
5	ANG36193	1

Assume that the following data record:

```
DG139201|Symphony No.9|BEETHOVEN|
```

is deleted from the data file and that the composer index is updated accordingly.

(continues in the next page...)

AFTER THIS DELETION, the composer index becomes:

A.

0	DVORAK	3
1	PROKOFIEV	5

0	********	*
1	LON2312	-1
2	ANG3795	4
3	COL31809	-1
4	RCA2626	-1
5	ANG36193	1

B.

0	BEETHOVEN	2
1	DVORAK	3
2	PROKOFIEV	5

0	********	*
1	LON2312	-1
2	ANG3795	0
3	COL31809	-1
4	RCA2626	-1
5	ANG36193	1

C.

0	BEETHOVEN	2
1	DVORAK	3
2	PROKOFIEV	5

0	********	*
1	LON2312	-1
2	ANG3795	-1
3	COL31809	-1
4	RCA2626	-1
5	ANG36193	1

D.

0	BEETHOVEN	2
1	DVORAK	3
2	PROKOFIEV	5

0	********	*
1	LON2312	-1
2	ANG3795	4
3	COL31809	-1
4	RCA2626	-1
5	ANG36193	1

E.

0	BEETHOVEN	1
1	DVORAK	3
2	PROKOFIEV	5

0	********	*
1	LON2312	-1
2	ANG3795	4
3	COL31809	-1
4	RCA2626	-1
5	ANG36193	1

... continued

15 Co-sequential Processing — 1 point

How many **read operations** and **key comparisons** are required in order to **match** the two following lists of letters, assuming that both lists are kept in files on disk:

```
List1: B, E, F, K, L, P, Z
List2: A, B, F, G
```

Important notes: Only count **read operations** in which a key (letter) is read; an attempt to read a key that detects the "end of file" should not be counted as a read operation.

A. number of read operations = 8   number of key comparisons = 5

B. number of read operations = 8   number of key comparisons = 8

C. number of read operations = 11   number of key comparisons = 8

D. number of read operations = 11   number of key comparisons = 5

E. number of read operations = 11   number of key comparisons = 28