Université d'Ottawa • University of Ottawa

Faculté de génie
École d'ingénierie
et de technologie de l'information

Faculty of Engineering
School of Information Technology
and Engineering

CSI 2131
FINAL EXAMINATION

Time : 3 Hours
Prof : H. Ural

This is a closed book exam.
Calculators are permitted
Give your answers on the questionnaire.

Mark allocation is as follows:

Q1	20
Q2.a	10
Q2.b	05
Q3.a	05
Q3.b	05
Q3.c	05
Q4.a	08
Q4.b	07
Q5	20
Q6	15

TOTAL 100

Student Name : _____

Student Number : _____

C.P. 450, Succ. A
Ottawa (Ontario) K1N 6N5 Canada

P.O. Box 450, Stn. A
Ottawa, Ontario K1N 6N5 Canada

(613) 562-5826 • Téléc./Fax (613) 562-5187

Q1. When sorting large files, a two-step process is applied: forming runs and merging runs.
Suppose that a file with 24 very large records with a blocking factor = 1 will be sorted in increasing order and that the input buffer in RAM can hold only 4 records. Assume that the keys for these 24 records are:
(Beginning of the file) (End of the file)
5, 16, 47, 12, 67, 21, 13, 17, 22, 3, 2, 9, 42, 20, 80, 14, 1, 11, 23, 53, 18, 63, 10, 4

While forming runs, replacement selection strategy is used and runs are written onto the disk with a blocking factor = 2 records, i.e., using an output buffer of 2 records.
While merging runs, a multi-phase merge strategy is used where an input buffer of 4 records and an output buffer of 2 records are used.
Show the steps of sorting the file and determine the required number of physical read and write operations at each step of the process and give the total at the end.

Q2. a) Encode the following sequence of characters using LZW algorithm:

ABCADCBCDACBADBCDA

@	STRING
0	A
1	B
2	C
3	D

BUFFER	INPUT	STRING	@	OUTPUT

b) What is the ratio of the length of the encoded sequence you obtained above with respect to the length of the minimum possible encoding of the original sequence?

Q3. Suppose that the following unsorted collection of fixed-length records are given:

Code	Hall	Dept
04	T	CSI
21	L	ENG
13	T	CSI
14	G	ENG
12	G	ACH
30	C	GEO
17	G	CSI
11	N	ACH
48	T	GEO
03	L	ENG

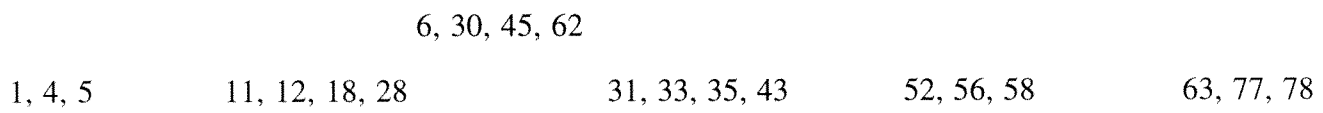
Assume that the primary key is Code which is to be used in accessing the records in skip sequential access.

a) Construct a simple index in RAM that is based on increasing order of keys which will allow binary search for finding a key and will allow skip sequential access to the record which has that key.

b) Construct a secondary index with respect to Hall and another secondary index with respect to Dept.

c) Convert the unsorted collection of records shown above into a multi-list file with embedded lists with respect to Hall and Dept. Give your answer in the box shown above.

Q4. Consider the B-tree of order 5 shown below.



a) Show the steps of inserting 29 into the tree.

b) Suppose that the tree shown above is a B*-tree of order 5. Show the steps of inserting 29 into the tree.

Q5. Consider a hash table called H_TABLE with N buckets where bucket size B = 1.
Consider also an overflow area called O_AREA with M buckets where bucket size B = 1.
Suppose that for collision resolution, external chaining is used such that the entries for all keys colliding with a key already stored in the hash table are stored in the overflow area organized and maintained outside of the hash table. Assume that an entry in the hash table or in the overflow area is declared as follows:
(KEY, RRN, P) where P is a pointer to a location in the overflow area. Note that (-1,-1,-1) represents an empty entry in the hash table or in the overflow area. In a hash table entry that is not empty, P = -1 indicates that no key has collided with the key in this entry. In an overflow area entry that is not empty, P = -1 indicates that the entry contains the last key in the list of synonyms.
Write a Pascal procedure called SEARCH that is defined as follows:
Given a HOME_ADDR and a KEY, it returns the RRN corresponding to the KEY if the KEY is found in H_TABLE or in O_AREA. Otherwise, it returns a FLAG set to TRUE indicating the KEY is not found.

Q6. Answer either part a) or part b), but not both.

a) For a B*-tree of order m and height h,

i) Derive an expression for the minimum total number of keys in the tree

ii) Derive an expression for the maximum total number of nodes in the tree

b) Suppose the following records with their corresponding keys will be stored in a file using the extendible hashing technique. The data file will be composed of pages that can be individually accessed and that each contains a maximum of two records.

Record #	key	hashed key
1	ax	1010101010101010
2	bs	0100010100001000
3	cx	0110010101010100
4	dy	0011001100110011
5	mn	0101101000100111
6	jg	1111111000011111
7	ol	0100111000011001

Show the steps of inserting the records in the data file and the resulting trie.