CSI2131, Winter 2001
Assignment 4
Due on Monday, April 9, 11:00am

The core assignment contains 4 questions adding up to 100 marks. You may also submit question 5 (optional), for a total mark of 115.

Question 1. [30 marks] B trees: insertions and deletions

V1) Variation using last key of a child (as in Chapter 9)
This is the standard way B trees were introduced in Chapter 9.

The format of a node is exemplified by the following node, which appeared in Figure 9.22 in page 406 of the textbook as the root of the B-tree:

	E		N		U		Z		-

V2) Variation using first key of a child.
This is a variation in which the first key of the child (rather than the last key) is stored in its parent.

The format of a node is exemplified by the following node, which shows how the root of the B-tree in Figure 9.22 would be modified to store the first key of its children:

A		K		P		V		-	

In this question you are going to operate on B trees of **order 3**.
Start with an empty tree. Insert the following keys into the B-tree:
AD, CA, ER, FA, FR, AY, BO.
Show your tree only after the underlined insertions.
Delete CA and show your tree.

You are going to do the above tasks in two different ways:

A) (15 marks) Employ variation **V1** and use **merging** when deleting.

B) (15 marks) Employ variation **V2** and use **redistribution** when deleting.

Question 2. [30 marks] **Simple prefix B+ trees: insertions and deletions**

Consider the simple prefix B+ tree of order 3 shown in Figure 10.7 in page 434 of the textbook. To avoid too much writing, you don't need to write all the names in the sequence set, and may only refer to block numbers.

Suppose a key added to block 5 results in a split of block 5 and the consequent addition of block 7, so blocks 5 and 7 are as follows:

$$\underset{\text{block 5}}{\boxed{\text{FABER-FINGER}}} \rightarrow \underset{\text{block 7}}{\boxed{\text{FINNEY-FOLK}}}$$

A) (10 marks) Show the tree after this insertion.

B) (15 marks) Suppose that, subsequent to the insertion, a deletion causes underflow and the consequent concatenation of blocks 4 and 5. What does the tree look like after the deletion?

C) (5 marks) Given the tree after the insertion in **part A.**, suppose that we need to redistribute by moving one name from block 5 to block 4, so that blocks 4 and 5 become:

$$\underset{\text{block 4}}{\boxed{\text{EMBRY-FABER}}} \rightarrow \underset{\text{block 5}}{\boxed{\text{FARCE-FINGER}}}$$

Show the tree after this redistribution.

Question 3. [25 marks] **Hashing: Predicting record distribution**

Please, show your work on the derivation of the following results. Suppose that 10,000 addresses are allocated to hold 8,000 records in a randomly hashed file and that each address can hold one record. Compute the following values:

A) (2 marks) the packing density for the file;

B) (5 marks) the expected number of addresses with no records assigned to them by the hash function;

C) (5 marks) the expected number of addresses with one record assigned to them (no synonyms);

D) (5 marks) the expected number of addresses with one record **plus** one or more synonyms;

E) (5 marks) the expected number of overflow records;

F) (3 marks) the expected percentage of overflow records.

Question 4. [15 marks] **Hashing: Progressive Overflow**

Consider the following Hash function:

```
Hash(String Key): Returns an Integer
{
Integer h
h = 2 * Key[0] + Key[1] + 3 * Key[2]
return h mod 6
}
```

A) (5 marks) What values do keys 'PAL' and 'LAP' map into? [A copy of the ASCII Table can be found in Appendix B of the textbook].

B) (5 marks) Given the following mappings (VAL, 3), (LAV, 1), (MAP, 5), (PAT, 3), (PET, 1) and (SET, 1), show the content of the Hash table of size 6 after

VAL, LAV, MAP, PAT, PET and SET

have been added in this order.
Notes: A Hash table of size 6 contains addresses 0-5. Assume that only one key can be held at each address and that Progressive Overflow (or Linear Probing) is used to resolve collisions. Remember that if the end of the table is reached when inserting a new key, then the add procedure wraps around the table, starting at position 0.

C) (5 marks) What is the average search length in this table? Show your work.

—————— Optional question ——————————

Question 5. [15 marks] **Hashing: Advanced Collision Resolution Methods**

A) (5 marks) Show the result of Chained Progressive Overflow after loading (MAP, 5), (PAT, 3), (PET, 1), (SET, 1), (SAT, 3), (CAT, 3) in this order in a Hash table of size 6, using two-pass loading.
Notes: A Hash table of size 6 contains addresses 0-5. Assume that only one key can be held at each address. Remember that if the end of the table is reached when inserting a new key, then the add procedure wraps around the table, starting at position 0.

B) (5 marks) What is the average search length in the hash table you just built?

C) (5 marks) Show the result of Chaining with a Separate Overflow Area after loading (MAP, 5), (PAT, 3), (PET, 1), (SET, 1), (SAT, 3), (CAT, 3) in this order.