COP 4720 - Information Systems and Databases II, Spring 2007,
Midterm

(This midterm consists of 6 questions in 3 pages.)
Good luck

Remarks:

• This midterm is out of 100 points. The value of each question/sub-question is written in
  square brackets, next to question number. You can get partial credits for your answers.
• Answer each question on a separate paper.
• Write your name on every paper you return.

Question 1 [6] Briefly describe the following terms

B. [2] Average seek cost
C. [2] Rotational latency

Question 2 [20]

A. [10] Briefly describe the TPMMS (Two Phase Multiway Merge Sort) algorithm.

B. [10] Assume that the size of a record is $2^{11}$ bytes and a 16 records fit in one block. Assume
  that the available memory size is 128 MB. Find the maximum number records that can be
  sorted using TPMMS.

Question 3 [14] Assume that a block can hold either two records or four key/pointer pairs or eight
  pointers. Assume that we have a database $D$ with 11 records with the following keys:

$$D = \{10, 10, 10, 15, 15, 20, 20, 30, 40, 40, 40\}$$


B. [8] Build a dense index with indirection when the data is placed on the disk in the following
  order.
Question 4 [18] Assume that a disk block can hold two records. Assume that we have a database \( D \) that contains initially three records with the following binary keys.

\[
D = \{01100, 00101, 11001\}
\]

Assume that linear hashing with at most 75% capacity is employed to index \( D \). Construct the index structure on \( D \). Insert two more records with keys 01010 and 11101 in this order. Show the index structure after the insertion of each record.

Question 5 [24] Assume that we have a database of gas stations \( G \) in the Santa Barbara area. For each gas station, we store its coordinates \((X, Y)\). Assume that \( G = (X, Y) \) consists of the following 11 records:

\[
G = \{(50, 100), (50, 120), (60, 80), (70, 50), (80, 45), (90, 80), (100, 30), (100, 150), (150, 45), (155, 200), (220, 150)\}
\]

The locations of the gas stations are shown in Figure 1. Assume that two records can fit in one block. Also assume that each dimension takes a value in the \([0, 256]\) interval.

Build a static-kd-tree, variation of kd-tree, on \( G \). Unlike the kd-tree, this tree blindly partitions a dimension into two from the middle of that dimension even if the points are not split evenly.

Question 6 [18] List at least one advantage of


B. [6] kd-tree over quad tree

C. [6] Extensible hashing over sparse index
Figure 1: Location of the gas stations on the X-Y plane