Exercises and hand-ins

Advanced database technology

February 20, 2003

Hand-in

To be handed in at the latest February 27 at 10.00 AM.

We have argued that bags can be more efficient that sets. This problem discusses some cases where bags may be less efficient. Assume in the following that a tuple fits in one block. Suppose that we have built a relation by simply adding N tuples one by one, not checking whether an added tuple was already present. Consider a sequence of queries that each require every tuple of the bag to be inspected.

1. Give an example of a bag of tuples for which these queries could take much longer than on the corresponding set of tuples.
2. Show that a bag may be turned into a set by sorting and performing one additional scan.
3. Suppose that whenever the number of I/Os spent on queries on a bag is equal to the number of I/Os needed to transform it into a set, we perform the transformation. Show that, no matter how many queries are performed, this strategy uses at most twice the number of I/Os compared to the best of:
   (a) Sorting before performing the queries.
   (b) Using a bag representation for all the queries.

Other exercises for discussion on February 27

1. Suppose we have a record with a fields that can be placed at any memory adress, b fields of 4 bytes that can start at an address that is a multiple of 4, and c fields of 8 bytes that can start at an address that is a multiple of 8. Show that it is possible to arrange the fields consecutively such that no space is wasted.
2. GUW 12.3.11 on page 589.
3. GUW 12.5.1 on page 601-2.
4. GUW 12.5.2 on page 602.