

## Introduction to Computer Science E15 – Discussion Sections – Week 44

1. Sequential files: Question 3 on page 446 and Problem 54 on page 454.
2. Merging: Question 1 on page 446.
3. Assume sets of numbers are represented by sequential files sorted on element value. For example, the set  $\{4, 7, 13, 9, 2\}$  is represented by a sequential file containing the sequence 2, 4, 7, 9, 13.

Describe algorithms for constructing  $A \cup B$  and  $(A \cup B) \cup C$  from  $A$ ,  $B$  and  $C$ . Note that  $(A \cup B) \cup C$  can be done by first computing  $A \cup B$  and computing the union of this with  $C$ . Instead of giving this solution, process the three files simultaneously, as you do with two files.

4. Assume the database relations  $A$  and  $B$  each are stored as sequential files of tuples, ordered according to attribute  $X$  (which is an attribute of both relations).

Sketch (details not necessary) an algorithm based on merging for executing the statement

$$C \leftarrow \text{JOIN } A \text{ and } B \text{ where } A.X = B.X$$

5. Assume again that the database relations  $A$  and  $B$  each are stored as sequential files, but now no longer ordered on the  $X$  attribute.

Describe an algorithm based on nested loops for executing the statement

$$C \leftarrow \text{JOIN } A \text{ and } B \text{ where } A.X = B.X$$

How many comparisons between tuples are performed (as a function of  $|A|$  and  $|B|$ , the numbers of tuples in each relations)?

Describe how to speed up the algorithm by first using hashing on each relation.

6. Do the problems on the binary search idea that were not done in week 40.