Binary I/O

11/14/2007





Opening Discussion

Do you have any questions about the quiz?
Let's look at solutions to the interclass problem.
Quantum computers: if and when.





Quicksort

- Quicksort can sort in place. At each step we pick a special element called the pivot. We move elements so that the pivot is in its proper place. Then we recursively call quicksort on the parts below and above the pivot.
- All the work happens going down the recursion.
- Expected behavior is O(n log n), but poor pivot selection leads to worst case of O(n²).
- For our code we will just take the first element as the pivot. We could easily pick another and swap it to the first spot.



- We have already played with files, but the files that we used were text files.
- Text files have the advantage that they are human readable and can be easily edited with any text editor.
- Most files that programs use a binary files. These are typically smaller and more efficient. With binary files we write the data to the disk as it appears in the memory of the machine.
 Binary files also allow us to do direct access which is vital in many applications.

Using Binary Files

- Working with binary files we use fopen and fclose just like normal, but you need to specify that it is a binary file in the fopen.
- Instead of fprintf or fscanf we use fread and fwrite.
- These functions directly copy from disk to memory or from memory to disk.
- Let's look at the man pages for these functions.



Random Access

- What really makes binary files significant is the ability to jump to any place in the file to read or write.
- This random access ability makes them very fast for accessing large amounts of data. Databases are built around this ability.
- The simplest form of direct access involves fixed length records. This is basically like having an array on disk.





Let's write some code to see how this works.





Minute Essay

- What are going to be some of the problems/challenges of dealing with random access files?
- Interclass Problem Make a binary file that has an array of integers. If you want a challenge, write a function to sort that which doesn't use an array, but sorts it on the file never loading more than two numbers at a time.