Review of Lectures 1-13

Fundamental File Processing Operations

- Physical file and logical file.
- Operations: open, close, read, write, seek.
- How these operations are implemented in C++.

Managing Files of Records

- Records, fields, keys, primary key, secondary keys.
- Field Structures: fixed length, variable length (with length indicator, with delimiters, or as "keyword=value")
- Record Structures: fixed length, variable length (with length indicator, with delimiter at the end, with fixed number of fields, using an index)
- Sequential search and direct access (by RRN).

Secondary Storage Devices

- Disks, tapes, CD-ROM.
- Characteristics, strengths and weaknesses of each of them.
- Disks: organization by sector and by block (blocking factor), tracks, surfaces, cylinders, cost of access (seek time, rotational delay, transfer time), clusters and extents.
- Tapes: data organization, blocks and gaps, space requirement (inches), nominal and effective recording density (bpi), nominal and effective transmission rate (bytes/sec)
- CD-ROM: lands and pits organization, differences with magnetic disks (CLV, CAV).

Basic I/O software and hardware, buffering

- Roles of operation system, I/O processor, disk controller.
- System I/O buffers. Buffering strategies: multiple buffering (double buffering, buffer pooling and pooling strategies).

Data Compression

- Techniques for data compression: compact notation, run-length encoding, Huffman code, Lempel-Ziv code.
- Huffman code: encoding, decoding, tree construction, number of bits required. [in Unix: pack and unpack]
- Lempel-Ziv: encoding, decoding, number of bits required. [in Unix: compress and uncompress]

Reclaiming space in files

- Strategies for record deletion: static: marking + storage compaction; dynamic: AVAIL LIST.
- AVAIL LIST with fixed-length records; AVAIL LIST with variable-length records and placement strategies: first-fit, best-fit, worst fit (their effect in fragmentation)

Sorting and Searching

- Sequential search, binary search.
- Internal sorting, external sorting.
- Keysorting

Indexing

- Indexing: primary index, secondary indexes.
- Addition, deletion and updates in primary and secondary indexes.

- Queries using combinations of keys.
- Inverted lists in secondary indexes.

Co-sequential processing

- Matching and merging.