CSci2111: Solution to Assignment # 3

1 Written Exercises

Question 1: Chapter 7, pp. 284-285

Problems 3: The status flag tells when the version of the index that is held in memory has been updated and not yet re-written to secondary storage. This means that the index that is on secondary storage is wrong and should not be used.

If a second program wants to use the index, then it has two choices:

- Wait until the status flag is reset (i.e., index on secondary storage is, again, valid).
- Share the correct index: the one stored in memory. [However, more coordination is needed and sometimes, the index in memory will be in the process of being revised and won't be fully accessible].

Problem 7: Possible Solution: Insert another Beethoven record and let the file processing software know somehow (e.g., by a link of some sort) that more than one index record has references to Beethoven records. Other solutions are also possible.

Problem 8:

- No change is made to the secondary index files.
- The following changes are made to the Label ID List file:
  - Append record 11 with label ID LON1259.
  - Insert record 11 into the 4th position in the list of Beethoven references by: 1) Changing the link field of record 5 (DG18807) to 11 and 2) Inserting the number "1" in the link field of record 11.

Question 2: Chapter 8, pp. 364-365

Problem 5: For a buffer size of $5,000$ **Kbytes**

- If $5,000$ Kbytes of memory are available for buffering, then a run can be expected to contain approximately $5,000,000/100 = 50{,}000$ records.
- At $50,000$ records per run, $8,000,000/50,000 = 160$ runs are needed to hold the $8,000,000$ record file.
- For the merge step, the $50,000$ record buffer must be divided into 160 parts, so 160 seeks will be required of each run . Since there are 160 runs, to complete the merge operation we end up making $160 * 160 = 25,600$ seeks.
- If the time taken for an average seek is 10 msec, then the total seek time required is $25,600 * 10 = 256,000$ msec or 256 seconds or 4 minutes and 16 seconds.

For a buffer size of $1,000$ Kbytes

- If $1,000$ Kbytes of memory are available for buffering, then a run can be expected to contain approximately $1,000,000/100 = 10{,}000$ records.
- At $10,000$ records per run, $8,000,000/10,000 = 800$ runs are needed to hold the $8,000,000$ record file.
- For the merge step, the $10,000$ record buffer must be divided into 800 parts, so 800 seeks will be required of each run . Since there are 800 runs, to complete the merge operation we end up making $800 * 800 = 640,000$ seeks.
- If the time taken for an average seek is 10 msec, then the total seek time required is $640,000 * 10 = 6,400,000$ msec or 6400 seconds or 1 hour and 46 minutes.

As the buffer size decreases, the time required to perform the merge operation becomes much much greater.

Problem 8:

- The file has r records divided into k runs and M is the amount of memory available.
- M/k is the amount of buffer space available per run.
- $M/(M/k) = k$ is the number of seeks needed to process one run.
- $k * k = k^2$ is the number of seeks required to process k runs.

If the length of a run is 2M (because of replacement selection)

- M/k is the amount of buffer space available per run.
- $2 * M/(M/k) = 2 * k$ is the number of seeks needed to process one run.
- $2 * k * k = 2 * k^2$ is the number of seeks required to process k runs (however, there will be fewer runs, so k will be smaller).