CS3310 – Kaminski - Fall 2012 World Data App 1.4
Asgn 4 Project Specs *external CodeIndex (as B Tree)*

OVERVIEW

This asgn is part of a larger project – but your task is to focus on the CodeIndex class. So, for A4 you only need to write:

A. **TempUserApp program** which only provides QueryByCode (QC) functionality (and NOT IN, LC, DC, QI, LI or DI).
   1. To simplify the development:
      a. handling of `MainData?.txt, TransData?.txt, LogSession.txt` is done right in the main program rather than in separate classes
      b. `MainData?.txt` is just a simple, easy-to-create temporary text file with several fields (with fake data), rather than the actual world data
      c. the auto testing feature (i.e., a for loop, with ? being 1,2,3) is handled in the main program rather than a separate program
   2. But ALL handling of the external index (in `CodeIndex?.bin` file) is handled within **CodeIndex class**, where the index is implemented as a B Tree.

B. TempSetupUtility program converts the 3 `CodeIndex?.txt` files into their equivalent `CodeIndex?.bin` files. It also displays the bin files for testing whether they're correct.

TempSetupUtility PROGRAM

The main algorithm:
1. open LogSession.txt file using truncate mode (rather than append mode)
2. for loop with fileNameSuffix (?) from 1 to 3
   - open the 2 files
   - convert CodeIndex?.**txt** file into CodeIndex?.**bin** file
   - go to the start of CodeIndex?.**bin** file
   - pretty-print it to LogSession.txt file
   - close the 2 CodeIndex files
3. close LogSession.txt file

3 CodeIndex?.txt FILES

Header record contains M (2 digits), RootPtr (3 digits), N (3 digits)
Other records contain:
    An array of M-1 triples – where each triple contains:
        TP (2 char, could be -1), KV (3 char), DRP (2 digits)
    The extra TP

Spaces between fields.
A <CR><LF> after each record.

3 CodeIndex?.bin FILES

Header Record contains 3 short's: M, RootPtr, N
Other records contain:
    An array of M TP's (short's)
    An array of M-1 KV's (char-arrays of size 3)
    An array of M-1 DRP's (short's)

NO spaces between fields.
NO <CR><LF> after each record.

LogSession.txt FILE
. . . . . *(part 1 - from TempSetupUtility)* . . . . . .

The pretty-printing of each of the **3 .bin files** looks like the following
(alter things appropriately based on M and fileNameSuffix):
[TP's and DRP's print as 2 columns so everything lines up!!!].

```
==============
CodeIndex1.bin     M is 5,  rootPtr is 11,  N is 22

----5 TP's----  ----4 KV's-----  --4 DRP's--
-1 -1 -1 -1 -1  ANN ANT APT ARE  03 04 19 01
-1 -1 -1 -1 -1  BOT BUG ]]] ]]]  05 02 00 00
. . .
```

Similarly for CodeIndex2.bin and CodeIndex3.bin

TempUserApp PROGRAM

SEE 4 IMPORTANT NOTES IN OVERVIEW

3 `TransData?.txt` FILES

Records contain: tranCode (QC), a space, code, <CR><LF>

3 `MainData?.txt` FILES

Just a text file with spaces between fields and <CR><LF>'s after each record.

Direct address on id with fixed-length records – so can do random access.

No Header record

Data records contain: id, code, restOfData

`LogSession.txt` FILE
. . . . . . ***(part 2 - from TempUserApp)*** **. . . . . .**

```
=======================
PROCESSING TransData 1

QC IMP
>> 10 IMP ish        49132
   [# nodes read:  1]
QC WMU
>> ERROR - no matching code in index
   [# nodes read:  3]
```

Similarly for TransData2 and TransData3

CodeIndex CLASS

Contains ALL handling of the external code index file.

Uses the BINARY versions of the file, NOT the TEXT versions.

Header record data is read in at start up.

Do NOT read in the entire FILE into MEMORY. This is an EXTERNAL storage structure, not an internal data structure.

This contains storage for a SINGLE node. There is NEVER more than ONE node in memory at once – so re-use the same storage space each time you read in a node into memory. (If you're using a separate node class, don't keep declaring new objects for every node you read – just keep re-using the same storage space).

To simplify searching by reducing the number of loop-stopping conditions to two, define the KV array to be of size M rather than M-1 (even though that's NOT what's in the file node), and initialize that extra KV as]]] (once at the start before reading in any nodes).

QueryByCode is a public method.

B Tree handling methods are private, including (with these names):
- ReadOneNode
- SearchOneNode

[NOTE: TRUTH IN ADVERTISING !!!]