Chan Ju Chong 정찬주 20150896

Setup Details

Coding Environment
• java version "1.8.0_121"
• Java(TM) SE Runtime Environment (build 1.8.0_121-b13)
• Java HotSpot(TM) 64-Bit Server VM (build 25.121-b13, mixed mode)

How to Launch Program(execute within folder)
> javac *.java
> java Main

Files
• data file is a csv file (data.csv)

📄 data
📄 Key.java
📄 Main.java
📄 Movie.java
📄 Node.java
📄 Tuple.java

Descriptions of data and B+ Tree

Source
• I used the information of the top 30 movies from the "Highest Rated IMDB 250 Titles" list.
• https://www.imdb.com/search/title?groups=top_250&sort=user_rating
• For each movie, I extracted the following information: movie title, year of release, rating, number of votes, name of director.

Schema
• Movie is the entity.
• It has 6 attributes: tid(int), movie(string), year of release(int), rating(float), no.of votes(int), director(string)

Number of Tuples
• My dataset consists of 30 tuples(movies)

B+ Tree
• Order of tree = 3
• I used the the pair (year of release, no.of votes) as the key of the B+ Tree index.
• The B+ tree is sorted in the ascending order of year of release and no. of votes. Year of release is compared first, then no. of votes . Examples: (2000, 1000) > (2000, 900), (2000, 900) > (1999,900)

B+ Tree Minor Details
• My B+ tree allows duplicate tid values to be stored.
• Uses linked list structure to store list of tids for each key of each leaf node.
• When there are 3 keys in a node, they are split so that the last 2 keys become another leaf node to the right.

Chan Ju Chong 정찬주 20150896

Instructions of Program

How to run the program
> javac *.java
> java Main

Testing Load
- Press 1 and enter.
- Specify beginning tid and press enter(tid begins at 1).
- Specify ending tid and press enter(there are 30 tuples so should not be bigger than this).
- To see if values have been added, press 2 and enter(for print).
- You can load multiple times if you want to load the same data.

Testing Print
- Press 2 and enter.
- If tree is empty, this will be shown.

```
Level 1: []
```
- If not, something like this will be shown(printing after loading tids 1~10):

```
Level 1: [(1993,1031413),(2003,1423464)]
Level 2: [(1966,592517),(1974,947932)], [(1994,1998810)], [(2008,1968220)]
Level 3: [((1957,559378),[8])] --> [((1966,592517),[9]),((1972,1369267),[2])] --
> [((1974,947932),[4])] --> [((1993,1031413),[6]),((1994,1560782),[5])] --> [((1
994,1998810),[1])] --> [((2003,1423464),[7])] --> [((2008,1968220),[3]),((2010,1
749014),[10])]
```

Testing Insert
- Press 3 and enter.
- Press tid between 1 and 30 and enter.(since we have 30 tuples)
- To see if value has been added, press 2 and enter(for print).
- Ex: insert tid 3, and print. (see result on right)

```
==============PRINT=============
Level 1: [((2008,1968220),[3])]
```

Testing Delete
- Press 4 and enter.
- Press tid between 1 and 30 and enter.(since we have 30 tuples)
   - You should insert tid first, otherwise you will get this message:
- To see if value has been deleted, press 2 and enter(for print).
- Ex: First insert tid 3. Then delete tid 3.(see result on right)

```
==============DELETE============
TUPLE ID: 2
Tuple #2 doesn't exist in the tree.

==============DELETE============
TUPLE ID: 3
Tuple #3 is deleted.
```

Testing Search
- Press 5 and enter.
- Enter a key to search for in the tree.
   - The key should be in the following format: (year of release,no.of votes)
- If you want to make sure a key is already in tree, insert tid 2 for example, and search for (1972,1369267).
- Result:
```
==============SEARCH===========
SEARCH KEY: (1972,1369267)
Found tuple IDs: [2]
Attributes: <tid, movie, year of release, rating, no.of votes, director>
Tuple #2:<2, "The Godfather", 1972, 9.2, 1369267, "Francis Ford Coppola">
```

Testing Range_Search
- Press 6 and enter.
- Enter two keys for a range to search for.
   - The two keys should be in the following format: [(year of release1, no.of votes1),(year of release2,no.of votes2)]
- For example, load tids 1~30, and search for [(1980,1004483),(1994,875630)]
- Result:
```
==========RANGE_SEARCH=========
SEARCH RANGE: [(1980,1004483),(1994,875630)]
Found pairs: [((1980,1004483),[14]),((1990,861577),[15]),((1991,1071252),[24]),(
(1993,1031413),[6]),((1994,875630),[23])]
Attributes: <tid, movie, year of release, rating, no.of votes, director>
Tuple #14:<14, "Star Wars: Episode V – The Empire Strikes Back", 1980, 8.8, 1004
483, "Irvin Kershner">
Tuple #15:<15, "Goodfellas", 1990, 8.7, 861577, "Martin Scorsese">
Tuple #24:<24, "The Silence of the Lambs", 1991, 8.6, 1071252, "Jonathan Demme">
Tuple #6:<6, "Schindler's List", 1993, 8.9, 1031413, "Steven Spielberg">
Tuple #23:<23, "Leon: The Professional", 1994, 8.6, 875630, "Luc Besson">
```