Project 2 Report

Jiang Miao   Fu Zhan

In this project we use tow method to achieve the indexing mechanism. The first way uses hashmap and the second way uses B plus tree. Each way include three interfaces
1 Put(string key, Number data_value) which for adding the index entry.
2 Get(string data_value) which for retrieving the key given the index.
3 Remove(string key) which for deleting the index.

1 Hashmap index mechanism

For hashmap, besides above three methods, class Createindex also includes writeFile(), readdbFile andreturnFile(). Read data from cs542.db by readdbFile() and extract key and value by readFile() and then put key/value pair into hashmap. When a specific value to be found, a correspondingkey can be decided and the position of this specific information in Movies.db can be found so that get all the movie information wanted.

In Class Indextest, two tests are created,

```java
System.out.println("Test one:");
String test1a = Createindex.get("1977|DVD");
Createindex.returnFile(test1a);
System.out.println("-----------------------------------------");
String test1b = Createindex.get("1990|VHS");
Createindex.returnFile(test1b);
System.out.println("-----------------------------------------");
String test1c = Createindex.get("2001|DVD");
Createindex.returnFile(test1c);
System.out.println("\n");

System.out.println("Test two:");
String test2a = Createindex.get("2000");
Createindex.returnFile(test2a);
System.out.println("-----------------------------------------");
String test2b = Createindex.get("2005");
Createindex.returnFile(test2b);
System.out.println("-----------------------------------------");
String test2c = Createindex.get("2010");
Createindex.returnFile(test2c);
```

And results are as follow picture shows:

```
Test one:
Not Found!
Wrong input!
-----------------------------------------
38
Pacific Heights,1990,VHS,Thriller,John Schlesinger,Daniel Pyne,USA,20th Century Fox,$15.00,0286
-----------------------------------------
1 4 14 15 19
Am??lie,2001,DVD,Comedy,Jean-Pierre Jeunet,Guillaume Laurant,France,Miramax,$15.00,0022
A Beautiful Mind,2001,DVD,Drama,Ron Howard,Akiva Goldsman,USA,Universal Pictures,$15.00,0054
Legally Blonde,2001,DVD,Comedy,Robert Luketic,Karen McCullah Lutz,USA,MGM,$15.00,0231
The Lord of the Rings: The Fellowship of the Ring,2001,DVD,Action,Peter Jackson,Phillippa Boyens,USA,New Line Cinema,$13.81,0707
Moulin Rouge,2001,DVD,Musical,Baz Luhrmann,Craig Pearce,USA,20th Century Fox,$15.00,0260


Test two:
6 8 16
Crouching Tiger Hidden Dragon,2000,DVD,Action,Ang Lee,Du Lu Wang,Taiwan,Asia / Columbia Pictures / Good Machine / Sony Pic,$15.00,0528
Gladiator,2000,DVD,Epic,Ridley Scott,John Logan,USA,DreamWorks / Universal  (Foreign),$15.00,0169
Malena,2000,DVD,Drama,Giuseppe Tornatore,Giuseppe Tornatore,Italy,Miramax,$15.00,0243
-----------------------------------------
20
Mr. & Mrs. Smith,2005,DVD,Action,Doug Liman,Simon Kinberg,USA,20th Century Fox,$15.00,0261
-----------------------------------------
Not Found!
Wrong input!
```

These results are all correct.

2 B plus tree index mechanism

For B+ tree, setting 3 assumptions:

BPlusTree Class Assumptions:

1. No duplicate keys inserted

2. Order D: D<=number of keys in a node <=2*D

3. All keys are non-negative

(a)Classes introduce

In BPlusTree Class get(), put() and remove() method are created. Search(), putHelper(), getHelper(), removeHelper and splitfNode() support above three methods and make sure they works fine. Use a search helper to locate the key recursively. Use a recursive insertHelper to locate the insert point, do insertion. Use a recursive deleteHelper to locate the delete point, do deletion. splitfNode()(including splitIndexNode and splitIndexNode) split an Node and return the new right node. Besides, HandleNodeUnderflow()(includes handleLeafNodeUnderflow and handleIndexNodeUnderflow) decides how to merge or redistribution when remove a key.

Class Utils includes methos assisting coding and tesing.

Class INdexNode incudes method which to make sure index remains sorted when inserting a new entry into this node.

Class LeafNode includes method which to make sure leaf remains sorted when inserting a new key/value pair into this node.

Class test include readTxtFile which to read data from data.db and store key and value into B+ tree.

(b)test result

Firstly all data should be inserted into B+ tree. In class utils, bulkInsert method put all key/value pairs into tree. And the result is shown:

```
 *  2/25/30/36/ *
 *  10/12/14/16/ * *  21/23/ * *  27/29/ * *  32/34/ * *  4/41/43/6/ *
[(0,1989|LaserDisc);(1,2001|DVD);]*[(10,2004|DVD);(11,1977|VHS);]*[(12,1977|VHS);(13,1997|DVD);]*[(14,2001|DVD);(15,2001|DVD);]*[(16,2000|DVD);(17,1977|VHS);(18,1999|DVD);(19,2001|DVD);]$[(
```

Degree is 3 and leaf nodes store 45 key/value pairs. The first line is root with 4 value and 5 pointers pointing to the second line(index node), they are the children of root. The third line is leaf node. For each node, it is separated by * from its siblings.

Root	* 2/25/30/36/ *
Index	* 10/12/14/16/ * * 21/23/ * * 27/29/ * * 32/34/ * * 4/41/43/6/ *
leaf	[(0,1989|LaserDisc);(1,2001|DVD);]*[(10,2004|DVD);(11,1977|VHS);]*[(12,1977|VHS);(13,1997|DVD);]*[(14,2001|DVD);(15,2001|DVD);]*[(16,2000|DVD);(17,1977|VHS);(18,1999|DVD);(19,2001|DVD);]$[(2,1999\|DVD);(20,2005\|DVD);]*[(21,1977\|VHS);(22,1994\|DVD);]*[(23,1994\|DVD);(24,1994\|DVD);]$[(25,1994|DVD);(26,1995|DVD);]*[(27,1994|DVD);(28,1995|DVD);]*[(29,1994|DVD);(3,1977|VHS);]$[(30,1995\|DVD);(31,1995\|DVD);]*[(32,1994\|DVD);(33,1994\|DVD);]*[(34,1994\|DVD);(35,1995\|DVD);]$[(36,1994|DVD);(37,1995|DVD);(38,1990|VHS);(39,1977|VHS);]*[(4,2001|DVD);(40,1998|DVD);]*[(41,2004|DVD);(42,1991|VHS);]*[(43,1987|VHS);(44,1989|LaserDisc);(5,1985|LaserDisc);]*[(6,2000|DVD);(7,1990|LaserDisc);(8,2000|DVD);(9,2004|DVD);]$

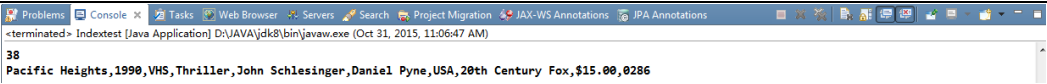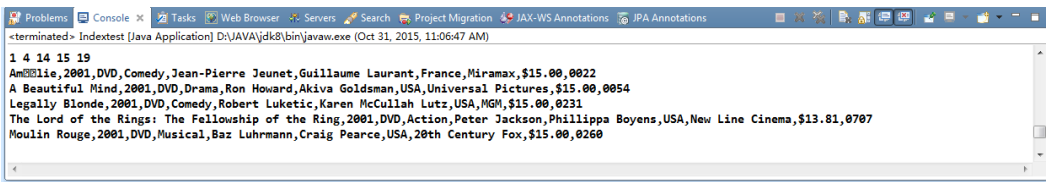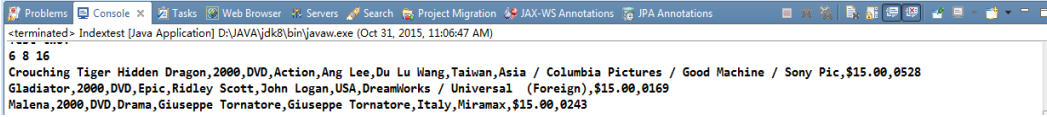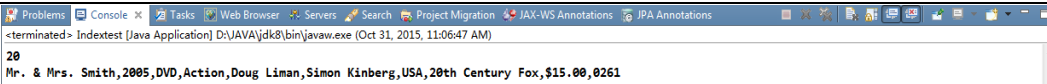Then after remove operation such as remove(0) remove(1),

```
 *  2/25/30/36/ *
 *  12/14/16/ * *  21/23/ * *  27/29/ * *  32/34/ * *  4/41/43/6/ *
[(10,2004|DVD);(11,1977|VHS);]*[(12,1977|VHS);(13,1997|DVD);]*[(14,2001|DVD);(15,2001|DVD);]*[(16,2000|DVD);(17,1977|VHS);(18,1999|DVD);(19,2001|DVD);]$[(2,1999|DVD);(20,2005|DVD);]*[(21,19
```

Root	* 2/25/30/36/ *
Index	* 12/14/16/ * * 21/23/ * * 27/29/ * * 32/34/ * * 4/41/43/6/ *
leaf	[(10,2004|DVD);(11,1977|VHS);]*[(12,1977|VHS);(13,1997|DVD);]*[(14,2001|DVD);(15,2001|DVD);]*[(16,2000|DVD);(17,1977|VHS);(18,1999|DVD);(19,2001|DVD);]$[(2,1999\|DVD);(20,2005\|DVD);]*[(21,1977\|VHS);(22,1994\|DVD);]*[(23,1994\|DVD);(24,1994\|DVD);]$[(25,1994|DVD);(26,1995|DVD);]*[(27,1994|DVD);(28,1995|DVD);]*[(29,1994|DVD);(3,1977|VHS);]$[(30,1995\|DVD);(31,1995\|DVD);]*[(32,1994\|DVD);(33,1994\|DVD);]*[(34,1994\|DVD);(35,1995\|DVD);]$[(36,1994|DVD);(37,1995|DVD);(38,1990|VHS);(39,1977|VHS);]*[(4,2001|DVD);(40,1998|DVD);]*[(41,2004|DVD);(42,1991|VHS);]*[(43,1987|VHS);(44,1989|LaserDisc);(5,1985|LaserDisc);]*[(6,2000|DVD);(7,1990|LaserDisc);(8,2000|DVD);(9,2004|DVD);]$

Result is right. When searching DVD movies made in 1977, VHS movies made in 1990, DVD movies made in 2001 I have to get search key and get responding value and then compare with 1997|DVD, 2001|VHS and 2001|DVD. It is a waste of time comparing with using hashmap. And a easy way to improve this is change key to a value related to year and format.

	Test 1
1	`Not Found!`

2	
	`38`
	`Pacific Heights,1990,VHS,Thriller,John Schlesinger,Daniel Pyne,USA,20th Century Fox,$15.00,0286`

3	
	`1 4 14 15 19`
	`Amélie,2001,DVD,Comedy,Jean-Pierre Jeunet,Guillaume Laurant,France,Miramax,$15.00,0022`
	`A Beautiful Mind,2001,DVD,Drama,Ron Howard,Akiva Goldsman,USA,Universal Pictures,$15.00,0054`
	`Legally Blonde,2001,DVD,Comedy,Robert Luketic,Karen McCullah Lutz,USA,MGM,$15.00,0231`
	`The Lord of the Rings: The Fellowship of the Ring,2001,DVD,Action,Peter Jackson,Phillippa Boyens,USA,New Line Cinema,$13.81,0707`
	`Moulin Rouge,2001,DVD,Musical,Baz Luhrmann,Craig Pearce,USA,20th Century Fox,$15.00,0260`

	Test 2

1	
	`6 8 16`
	`Crouching Tiger Hidden Dragon,2000,DVD,Action,Ang Lee,Du Lu Wang,Taiwan,Asia / Columbia Pictures / Good Machine / Sony Pic,$15.00,0528`
	`Gladiator,2000,DVD,Epic,Ridley Scott,John Logan,USA,DreamWorks / Universal  (Foreign),$15.00,0169`
	`Malena,2000,DVD,Drama,Giuseppe Tornatore,Giuseppe Tornatore,Italy,Miramax,$15.00,0243`

2	
	`20`
	`Mr. & Mrs. Smith,2005,DVD,Action,Doug Liman,Simon Kinberg,USA,20th Century Fox,$15.00,0261`

3	`Not Found!`