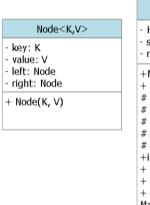
BLM19202E Data Structures Programming Assignment #3 Movie Recommendation System using Heap-Based Collaborative Filtering

HAMZA ALAHMEDİ- 2021221358 / RAMİ EYÜPOĞLU-1921221355

1- Design:



MaxHeap		
- Heap: Node[] - size: int		
- maxsize: int		
+MaxHeap(int) + isEmpty(): boolean # parent(int) # leftChild(int) # rightChild(int) # isLeaf(int) # swap(int, int) +insert(Node <k, v="">) + print() + extractMax() + getMaxUsers(numUser) MaxHeap): MaxHeap</k,>	#maxHeapify(int) rs: int, heap:	

reader

+readExcel (String): int[][]
+ printMatrix(int[][])
+ cosineSimilarity(int[], int[]):
double
+ printMaxUsers(int, MaxHeap)

- model: DefaultListModel<String>
- targetUser: String
- users: String
- movieFile: String
- movieFile: String
- movieHeap: MaxHeap<Integer, String>

+getHighestRatedMovies(matrix: int[][], numHighestRatedMovies: int): List<Integer>
+printList(list: List<T>): void
+createNewMatrix(originalMatrix: int[][], userHeap: MaxHeap<Double, Integer>): int[][]
+readMovieFile(filename: String): MaxHeap<Integer, String>
+searchMovieIDs(movieIDs: List<Integer>, movieHeap: MaxHeap<Integer, String>): List<String>
+getSelectedMovies(comboBoxes: List<JComboBox<String>): List<Object>
+searchMovieIDsyNames(movieNames: List<Object>, movieHeap: MaxHeap<Integer, String>): List<Integer>
-jButton1ActionPerformed(evt: ActionEvent): void
-jButton2ActionPerformed(evt: ActionEvent): void

Pseudo-code and algorthim steps:

a. First Window:

- 1- Read the target_user file and main_data file and create a matrix for each file: Use the readExcel method in the Reader class to read files and obtain a matrix representation of the data.
- 2- Calculate the cosine similarity: Iterate over each user in the main data matrix. For each user, calculate the cosine similarity with the target user by using the cosineSimilarity method in the Reader class. Store the user ID and cosine similarity value as a pair.
- 3- Add user IDs with cosine similarity to a multi-value MaxHeap: Create a new instance of the MaxHeap class with the appropriate data types. Add the user IDs and cosine similarity values as nodes to the heap.

- 4- Get the x most similar users: Use the extractMax method in the MaxHeap class to extract the x highest cosine similarity users from the heap. Create a new instance of the MaxHeap class specifically for the similar users and add the extracted nodes to this new heap.
- 5- Read the main data file and create a new matrix: Iterate over each user ID in the similar users heap. For each user ID, retrieve the entire row of data from the main data matrix corresponding to that user. Create a new matrix with these rows.
- 6- Get the K highest rated movies IDs: Determine the K highest rated movies by implementing the appropriate logic. Create a list to store the movie IDs.
- 7- Read the movies file and create a MaxHeap with movie ID and name: Use the readExcel method in the Reader class to read the movies file and obtain a matrix representation of the data. Create a new instance of the MaxHeap class to store the movie IDs and names.
- 8- Get the movie names from the MaxHeap: Iterate over the movie IDs in the list of K highest rated movies. Search for each movie ID in the movie heap and retrieve the corresponding movie name. Add the movie names to a JList or any suitable data structure.

b. Second Window:

- 1- Obtain the newUserId from a text field.
- 2- Retrieve the selected movie names from combo boxes.
- 3- Collect the movie ratings from text fields.
- 4- Search the movie heap for the movie IDs corresponding to the selected movie names.
- 5- Create a new vector to hold the newUserId and the movie ratings, with each rating stored at the index corresponding to its movie ID.
- 6- Calculate the cosine similarity between the new user and all other users in the system.
- 7- Add the user IDs along with their cosine similarity values to a multi-value MaxHeap data structure.
- 8- Retrieve the x most similar users by extracting the highest cosine similarity users from the MaxHeap.
- 9- Read the main data file and create a new matrix containing the data of the similar users.
- 10- Determine the K highest rated movie IDs by implementing the appropriate logic.
- 11- Read the movies file and create a MaxHeap data structure to store movie IDs and names.
- 12- Retrieve the movie names from the MaxHeap by searching for each movie ID in the list of highest rated movies.
- 13- Populate a JList or suitable data structure with the recommended movie names.

implementation details:

```
public class Node<K extends Comparable<K>, V> {
    K key; // the cossinSimilraty
    V value; // the User id
    Node<K, V> left, right;

public Node(K key, V value) {
    this.key = key;
    this.value = value;
    left = right = null;
    }
}
```

The Node class(Generic type) has the following attributes:

key: Represents the cosine similarity, value: Represents the user ID.

left: Represents the left child node, right: Represents the right child node.

The class provides a constructor that takes a key and a value as parameters and initializes the attributes accordingly. The left and right child nodes are initially set to null.

Reader class:

```
java.io.FileInputStream;
          import java.io.IOException;
         import java.util.ArrayList;
         import java.util.Scanner;
14
15
              public static int[][] readExcel(String filePath) {
16
17
                    int[][] matrix = null;
                         File file = new File(filePath);
FileInputStream fis = new FileInputStream(file);
Scanner scanner = new Scanner(fis);
20
21
22
24
25
                         while (scanner.hasNextLine()) {
                              String line = scanner.nextLine();
29
30
                         matrix = new int[numRows][numCols];
37
38
                              for (int j = 0; j < numCols; j++) {
   matrix[i][j] = Integer.parseInt(values[j]);</pre>
42
43
45
94
47
48
                         e.printStackTrace();
```

readExcel(String filePath): This method takes a file path as input and reads an Excel file. It reads the contents of the file and stores them in a 2D integer array (matrix) by splitting each line at commas. It returns the resulting matrix.

```
public static void printMatrix(int[][] matrix) {
    for (int[] row : matrix) {
        for (int element : row) {
            System...print(element + "\t"); // Use "\t" for tab-separated columns
    }
    System...println(); // Move to the next line after printing each row
}

public static void printMaxUsers(int numUsers, MaxHeap heap) {
    for (int i = 0; i < numUsers; i++) {
        Node<Double, Integer> maxNode = heap.extractMax();
        if (maxNode == null) {
            break; // If there are no more nodes in the heap, exit the loop
        }
        System...println("User ID: " + maxNode... + ", Value: " + maxNode...);
}
```

printMatrix(int[][] matrix): This method takes a 2D integer array (matrix) as input and prints its contents. It iterates over each row and column of the matrix and prints each element, separated by a tab (\t). After printing each row, it moves to the next line.

printMaxUsers(int numUsers, MaxHeap heap): This method takes the number of users to print (numUsers) and a MaxHeap object (heap) as input. It iterates numUsers times and extracts the maximum node from the heap. If there are no more nodes in the heap, the loop is exited. It then prints the user ID and value of the extracted node.

```
// Method to calculate the cosine similarity
public static double cosineSimilarity(int[] vectorA, int[] vectorB) {
    double dotProduct = 0.0;
    double normA = 0.0;
    double normB = 0.0;
    for (int i = 0; i < vectorA.length; i++) {
        dotProduct += vectorA[i] * vectorB[i];
        normA += Math.pow(vectorA[i], 2);
        normB += Math.pow(vectorB[i], 2);
}
return dotProduct / (Math.sqrt(normA) * Math.sqrt(normB));
}</pre>
```

cosineSimilarity(int[] vectorA, int[] vectorB): This method takes two integer arrays (vectorA and vectorB) as input and calculates the cosine similarity between them. It calculates the dot product, as well as the norms of the vectors, and then divides the dot product by the product of the vector norms. The resulting cosine similarity value is returned.

MaxHeap class:

Node<K extends Comparable<K>, V>: This is a nested class representing a node in the heap. It has a key of type K (which must be comparable) and a value of type V.

MaxHeap(int capacity): This is the constructor of the MaxHeap class. It initializes the heap with a given capacity. It creates an array Heap of Node objects with the specified capacity, sets the initial size to 0, and stores the maximum capacity.

isEmpty(): This method checks if the heap is empty by comparing the size with 0. It returns true if the heap is empty, and false otherwise.

parent(int pos): This private method calculates the index of the parent node given a position pos in the heap.

leftChild(int pos): This private method calculates the index of the left child node given a position pos in the heap.

rightChild(int pos): This private method calculates the index of the right child node given a position pos in the heap.

isLeaf(int pos): This private method checks if the node at the given position pos is a leaf node. It returns true if the node is a leaf, and false otherwise.

```
// Swaps the nodes at the given positions in the Heap array
private void swap(int fpos, int spos) {
    Node(K, V) tmp;
    tmp = int[pos];
    int [fpos] = int[spos];
    int [spos] = tmp;
}

// Maintains the max heap property starting from the given position
// Maintains the max heap property starting from the given position
// Maintains the max heap property starting from the given position
private void maxHeapify(int pos) {
    int largest = pos;
    int left = leftChild(pos);
    int right = rightChild(pos);

// Compare the left child with the current position
if (left < int && int [left]. ...compareTo(int [largest]. ) > 0) {
        largest = left;
    }

// Compare the right child with the current largest
if (right < i& int [right]. ...compareTo(int [largest]. ) > 0) {
        largest = right;
    }

// If the largest element is not the current position, swap them
if (largest != pos) {
        swap(pos, largest);
        maxHeapify(largest);
    }
}
```

swap(int fpos, int spos): This private method swaps the nodes at the given positions fpos and spos in the Heap array.

maxHeapify(int pos): This private method maintains the max heap property starting from the given position pos. It compares the node at pos with its children, and if necessary, swaps them to maintain the max heap property. It recursively calls itself on the child nodes.

```
// Inserts a new node into the max heap
public void insert(Node<K, V> node) {
    if (new >= new node) {
        throw new IllegalStateException("Heap is full. Cannot insert more elements.");
    }

    int current = node;
    int current = node;
    int current = node;
    int current > 0 && Node into [current]. Node compareTo(Node [parent(current)]. Node of the second parent(current)]. Node of the second parent (node into the maximum three nodes property maxHeapify(0);
    if (leftChild(0) < node) {
        maxHeapify(leftChild(0));
    }
    if (rightChild(0) < node) {
        maxHeapify(rightChild(0));
    }
}</pre>
```

insert(Node<K, V> node): This method inserts a new node into the max heap. It checks if the heap is already full and throws an exception if so. It adds the node to the Heap array at the current size position and increments the size. Then, it bubbles up the inserted node by comparing it with its parent and swapping them if necessary. Finally, it performs maxHeapify on the root and its children to ensure the maximum heap property is maintained.

PrintHelper(int index): This private recursive method is called by Print() to traverse the heap in a post-order traversal. It prints the nodes by recursively traversing the left and right subtrees and then printing the current node.

Print(): This method prints the nodes in the heap in a post-order traversal. It calls the PrintHelper method to traverse the left subtree, then the right subtree, and finally prints the current node.

```
// Extracts the maximum node from the max heap
public Node<K, V> extractMax() {
    Node<K, V> popped = Heap[0];
    Heap[0] = Heap[--rive];
    maxHeapify(0);
    return popped;
}
```

extractMax(): This method extracts the maximum node from the max heap. It removes the root node (which contains the maximum value) and replaces it with the last node in the heap. It then calls maxHeapify(0) to restore the max heap property and returns the extracted node.

```
// Retrieve the maximum users from the heap and return a new MaxHeap object
public static MaxHeap<Double, Integer> getMaxUsers(int numUsers, MaxHeap heap) {
    MaxHeap<Double, Integer> maxUserheap = new MaxHeap<>(numUsers);
    for (int i = 0; i < numUsers; i++) {
        Node<Double, Integer> maxNode = heap.extractMax();
        if (maxNode == null) {
            break; // If there are no more nodes in the heap, exit the loop
        }
        maxUserheap.insert(maxNode);
}

return maxUserheap;
}
```

getMaxUsers(int numUsers, MaxHeap heap): This static method retrieves the maximum users from the heap and returns a new MaxHeap object. It extracts the maximum node numUsers times from

the given heap and inserts them into a new MaxHeap called maxUserHeap. It returns maxUserHeap containing the maximum numUsers nodes.

GUI class:

```
43
44
                   BufferedReader br = new BufferedReader(new FileReader(targetUser));
                   String line;
48
                   int lineNumber = 0;
49
.50
                       lineNumber++:
.52
                       if (columns.length > 0) {
.53
                           String columnValue = columns[0]; // Get the value of the first column
.54
.55
.56
                           System. ..... println("Invalid column index on line " + lineNumber);
.58
.59
                  br.close();
              } catch (IOException e) {
                  System.out.println(e);
                           setSelectedIndex(1);
```

this code reads a CSV file, extracts values from the first column, and populates a JComboBox with those values, while also handling error cases where the column index is invalid or an IO exception occurs during file reading.

```
public static List<Integer> getHighestRatedMovies(int[][] matrix, int numHighestRatedMovies) {
    List<Integer> highestRatedMovieIndices = new ArrayList<>();

    for (int user = 0; user < matrix. user!) {
        int[] ratings = matrix[user];

        List<Integer> userHighestRatedMovies = new ArrayList<>();

for (int i = 0; i < numHighestRatedMovies; i++) {
        int maxRatingIndex = -1;
        int maxRating = 0;

for (int movie = 1; movie < ratings. user; movie++) {
            int rating = ratings[movie];

            if (rating > maxRating & ( luserHighestRatedMovies.contains(movie)) {
                maxRatingIndex = movie;
            }

            if (maxRatingIndex = movie;
            }

            if (maxRatingIndex i= -1) {
                 userHighestRatedMovies.add(maxRatingIndex);
            }

            highestRatedMovieIndices.addAll(userHighestRatedMovies);
            }

            return highestRatedMovieIndices;
}
```

getHighestRatedMovies: The given method takes a two-dimensional array (matrix) representing movie ratings by users, and an integer (numHighestRatedMovies) specifying the number of highest-rated movies to be returned.

The code finds the indices of the highest-rated movies across all users based on a given movie rating matrix. It returns a list of these indices, limited to a specified number of highest-rated movies per user.

```
public static int[][] createNewMatrix(int[][] originalMatrix, MaxHeap<Double, Integer> userHeap) {
    int numRows = userHeap.com;
    int numCols = originalMatrix[0].length;
    int[][] newMatrix = new int[numRows][numCols];

int rowIndex = 0;
    while (luserHeap.isEmpty()) {
        int userId = userHeap.extractMax().com;
        for (int col = 0; col < numCols; col++) {
            newMatrix[rowIndex][col] = originalMatrix[userId][col];
        }
        rowIndex++;
    }

return newMatrix;
}
</pre>
```

creatNewMatrix: This method creates a new matrix based on an original matrix and a max heap of user IDs. It extracts user IDs from the max heap in descending order of their priority (max value) and assigns the corresponding row from the original matrix to the new matrix. The new matrix will have a number of rows equal to the size of the max heap and the same number of columns as the original matrix. The resulting new matrix represents a rearrangement of rows from the original matrix based on the priority of user IDs in the max heap.

```
public static MaxHeap<Integer, String> readMovieFile(String filename) {
86
               MaxHeap<Integer, String> movieHeap = new MaxHeap<>(10000);
               try (BufferedReader reader = new BufferedReader(new FileReader(filename))) {
89
                   String line;
90
                       String[] parts = line.split(",", 3);
                       if (parts.)
                            int movieId = Integer.parseInt(parts[0].trim());
94
                           Node<Integer, String> node = new Node<>(movieId, movieName);
96
                           movieHeap.insert(node);
98
99
                 catch (IOException e) {
                   e.printStackTrace();
101
102
103
               return movieHeap;
104
```

readMovieFile(String filename): This method reads a movie file specified by the filename parameter and creates a max heap (MaxHeap<Integer, String>) based on the data in the file.

The method uses a BufferedReader to read the file line by line. Each line is split into three parts using a comma as the delimiter. If a line contains three parts, indicating a valid movie entry, the method extracts the movie ID (parsed as an integer) and the movie name from the parts array.

Then, a Node<Integer, String> object is created using the movie ID and movie name, and this node is inserted into the movieHeap max heap. The max heap is initialized with a capacity of 10000. After processing all the lines in the file, the method returns the populated movieHeap max heap containing the movie IDs and names read from the file.

This method searches for movie names corresponding to a given list of movie IDs within a movieHeap max heap. it initializes an empty list movieNames to store the found movie names and an empty list extractedNodes to store the extracted nodes from the movieHeap.

Next, it extracts all nodes from the movieHeap using the extractMax() method and adds them to the extractedNodes list .Then, for each movie ID in the movieIDs list, it searches for a matching movie ID within the extractedNodes. If a match is found, it retrieves the corresponding movie name from the node and adds it to the movieNames list. If no match is found, it prints a message indicating that the movie ID was not found and adds the string "Movie Not Found" to the movieNames list.

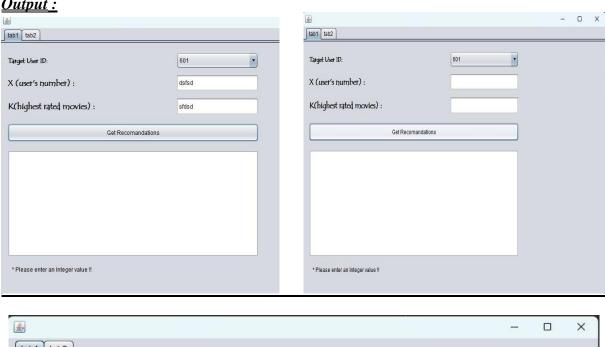
Finally, it returns the movieNames list, which contains the names of the movies corresponding to the provided movie IDs, or "Movie Not Found" for IDs that were not found in the movieHeap

```
d jButton1ActionPerformed(java.awt.event.ActionEvent =
                   String xValue = "......getText();
String kValue = "......getText();
314
315
316
317
318
                   319
320
321
322
323
324
                        int[][] main data matrix = Reader.readExcel()
327
328
331
                                                                                            h; user id++) {
332
333
                           double similarity = Reader.cosineSimilarity(target_user_matrix[
Node<Double, Integer> node = new Node<>(similarity, user_id);
334
335
336
337
                        int numberOfUsers = Integer.parseInt(xValue);
338
339
                        MaxHeap<Double, Integer> maxUserHeap = MaxHeap.getMaxUsers(numberOfUsers, maxHeap);
340
343
344
                        printList (movieIds);
345
346
                        List<String> movieNames = searchMovieIDs(movieIds, movieHeap);
printList(movieNames);
347
348
349
350
                        clear();
for (String movie : movieNames) {
351
352
                                    .addElement(movie);
```

It checks if the values of User and Movie text fields are positive integer values using regular expressions. If any of them is not a positive integer, it sets the text of jLabel1 to indicate that the user should enter an integer value.

If both User and Movie values are positive integers the method performs various operations including reading data from files, calculating similarity, creating and manipulating max heaps, retrieving movie information, and updating the GUI with the results based on user input.

Output:



		177	×
tab1 tab2			
Target User ID:	602		
X (user's number) :	2		
K(highest rated movies) :	5		
Get Recoma	indations		
Love Affair (1994)			
2001: A Space Odyssey (1968)			
A CONTRACTOR OF THE CONTRACTOR			
Angel Baby (1995)			
Angel Baby (1995) Fallen Angels (Duo luo tian shi) (1995)			
Fallen Angels (Duo luo tian shi) (1995)			
Fallen Angels (Duo luo tian shi) (1995) Son of Flubber (1963)			
Fallen Angels (Duo luo tian shi) (1995) Son of Flubber (1963) Blue in the Face (1995)			
Fallen Angels (Duo luo tian shi) (1995) Son of Flubber (1963) Blue in the Face (1995) Spellbound (1945)			

Second panel:

```
BufferedReader br = new BufferedReader(new FileReader(new FileRead
                                                                                                                         String line;
Random random = new Random();
249
250
 252
 254
 256
 258
259
                                                                                                                                                                                            comboBox = |ComboBox2;
 260
261
 262
263
264
265
                                                                                                                                                                                               comboBox =
 267
268
 269
270
 271
273
274
275
276
                                                                                                                                                                      int numMovies = 0;
while (numMovies <</pre>
                                                                                                                                                                                                            (numMovies < 10 && (line = br.readLine()) != null) {
 278
                                                                                                                                                                                       String[] columns = line.split(",");
if (columns. > 1) {
 279
 280
281
                                                                                                                                                                                                                       String movieName = columns[1];
282
283
                                                                                                                                                                                                                       comboBox.addItem(movieName);
                                                                                                                                                                                                                    numMovies++;
284
285
 289
                                                                                                                      br.close();
291
292
```

The code reads a movie file, skips the header line, and randomly selects movies to populate several JComboBox components. Each JComboBox is associated with a specific number (2 to 6), and the code adds 10 movies from the file to each corresponding JComboBox. The goal is to populate the JComboBox components with a selection of movies for user interaction or selection purposes.

```
public static List<Integer> searchMovieIDsByNames(List<Object> movieNames, MaxHeap<Integer, String> movieNeap)
List<Integer> movieID = new ArrayList<>();
List<Node<Integer, String> extractedNodes = new ArrayList<>();

// Extract all nodes from the movie heap
while (ImovieHeap,isEmpty()) (
extractedNodes.add(movieHeap.extractMax());
}

// Search for movie IDs in the extracted nodes
for (Object movieName : movieNames) (
boolean found = false;
for (Node<Integer, String> node : extractedNodes) (
if (node. ....equals(movieName)) (
int movieid = node. ;
movieID.add(movieid);
found = true;
break;

)

if (Ifound) (
System. ...println("Movie name " + movieName + " not found in the movie heap.");
movieNames.add("Movie Not Found");
}

return movieID;
}

return movieID;
}
```

```
int[] ratings = getValuesFromTextFields(textFields);
int[] newUserMatrix = createNewUserMatrix(ratings, movieIds, newUserId, main_data_matrix[0].
742
743
744
745
746
747
748
749
750
751
                        MaxHeap<Double, Integer> maxHeap = new MaxHeap<>(10000);
for (int user_id = 1; user_id < main_data_matrix. ; user_id++) {
    double similarity = Reader.cosineSimilarity(newUserMatrix, main_data_matrix[user_id]);
    Node<Double, Integer> node = new Node<>(similarity, user_id);
752
753
754
755
                              maxHeap.insert(node);
                        756
757
758
759
760
761
762
763
                        int maxUserMatrix[][] = createNewMatrix(main_data_matrix, maxUserHeap);
764
765
766
767
                        List<String> movieNames = searchMovieIDs(highestRatedmovieIds, movieHeap2);
printList(movieNames);
768
769
                                    .addElement(movie);
770
771
772
773
                             st2.setModel(model);
774
775
             private void jButton2ActionPerformed(java.awt.event.ActionEvent evt) {
   boolean jsWork = true;
706
                        .getText().isEmpty() || .getText().isEmpty() ||
|| .getText().isEmpty() || .getText().isEmpty
|| .getText().isEmpty() || .getText().isEmpty()) (

isWork = false;
707
708
                   ) else (
                        string regex = "[1-5]";
721
722
723
                        .setText("");

System. .println("test test");

MaxHeap<fractring> moyieHeap = readMoyieFile(
List<JComboBox<String> comboBoxes = new ArrayList<>();
comboBoxes.add(
);
                        List<Object> selectedMovies = qetSelectedMovies(comboBoxes);
```

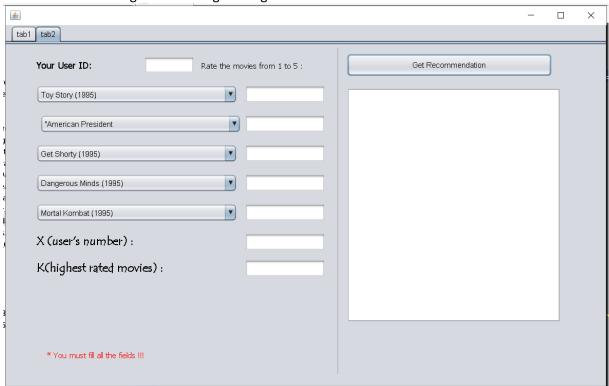
The jButton2ActionPerformed method is triggered when a button is clicked. It performs several actions to provide movie recommendations based on user input. First, it checks if certain fields are empty and displays an error message if they are. Then, it reads data from Excel files to create a target user matrix and a main data matrix, which represent user preferences and movie ratings. Next, it retrieves the selected movies from combo boxes and converts them into movie IDs. It also collects ratings provided by the user from text fields and creates a new user matrix using these ratings and the selected movie IDs.

To find similar users, the method calculates the similarity between the new user matrix and each user in the main data matrix using cosine similarity. It stores the similarity scores in a max heap. After that, it extracts the highest rated movies from the users with the most similar profiles. It searches for the corresponding movie names using the movie IDs.

Finally, it populates a model with the movie names and updates a list component (jList2) in the user interface with the recommended movies. This allows the user to view and select from the recommended movie list.

OutPut:

Control the field and gave a warning message .



This is the final output.

