Visvesvaraya Technological University Belagavi-590018, Karnataka



A Mini Project Report on

"INDEXING FOR LIBRARY DATA"

Submitted in partial fulfilment of the requirement for the File Structures Laboratory with Mini Project [17ISL68]

Bachelor of Engineering

in

Information Science and Engineering

Submitted by AMULYA.R[1JT17IS004]

Guided by Asst.Prof VADIRAJA.A



Department of Information Science and Engineering Jyothy Institute of Technology Tataguni, Bengaluru-560082

2019-20

Jyothy Institute of Technology Tataguni, Bengaluru-560082 Department of Information Science and Engineering



CERTIFICATE

Certified that the mini project work entitled "INDEXING FOR LIBRARY DATA" carried out by AMULYA.R [1JT17IS004] bonafide student of Jyothy Institute of Technology, in partial fulfilment for the award of Bachelor of Engineering in Information Science and Engineering department of the Visvesvaraya Technological University, Belagavi during the year 2019-2020. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the Report deposited in the departmental library. The mini project report has been approved as it satisfies the academic requirements in respect of Mini Project work prescribed for the said Degree.

Vadiraja.A Guide, Asst.Professor Dept. Of ISE Dr. Harshavardan Tiwari Professor & HoD Dept. Of ISE

External Viva Examiner

1.

2

Signature with Date:

ACKNOWLEDGEMENT

Firstly, I am very grateful to this esteemed institution **Jyothy Institute of Technology** for providing me an opportunity to complete my project.

I express my sincere thanks to my Principal **Dr. Gopalakrishna K** for providing me with adequate facilities to undertake this project.

I would like to thank **Dr. Harshwardhan Tiwari, Professor and Head of Information Science and Engineering Department** for providing for his valuable support.

I would like to thank my guide **Mr Vadiraja.A**, **Asst. Prof.** for his keen interest and guidance in preparing this work.

Finally, I would thank all my friends who have helped me directly or indirectly in this project.

AMULYA.R [1JT17IS004]

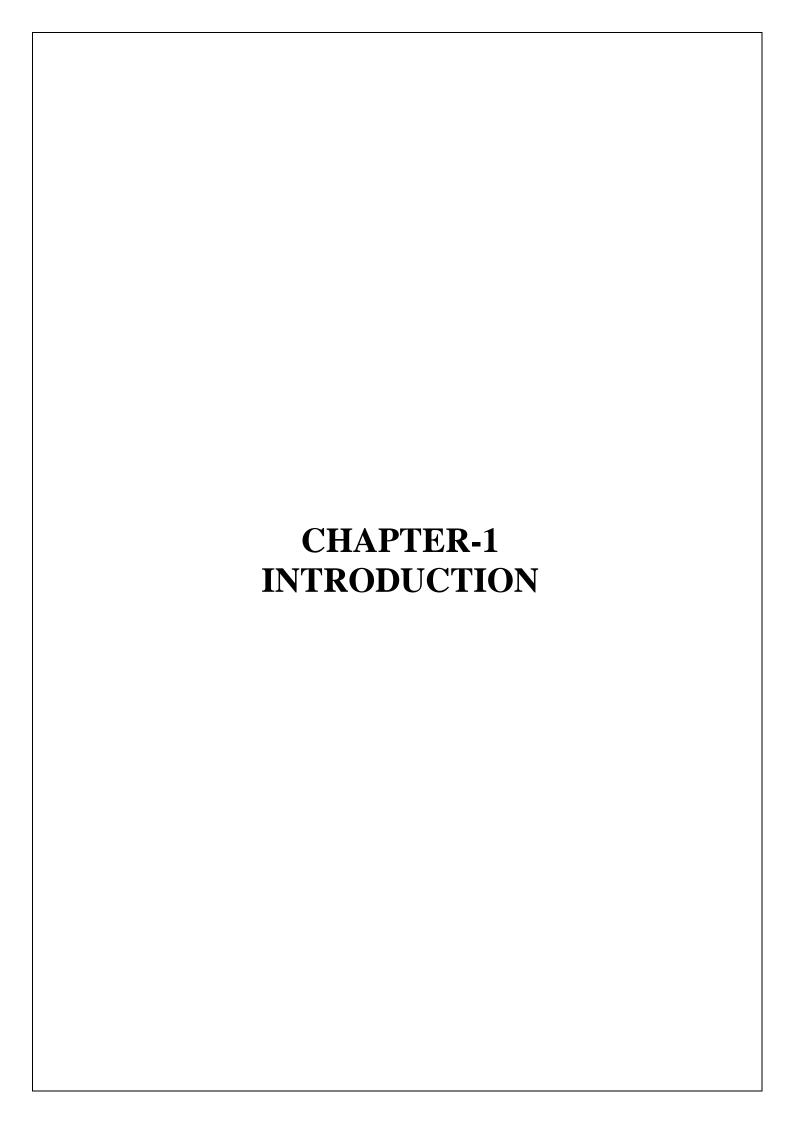
ABSTRACT

The project will be implemented on Eclipse platform, with the operating system WIINDOWS. The purpose of this project is to insert, delete, search and modify a required data-set in a large data and gathering the information corresponding to the ky value. Depending on the search parameter namely primary/secondary index a uniquely generated or a common set of data would be retrieved. The required software and hardware are easily available and easy to work with.

In this Library management system, we can enter the record of new records of new books and retrieve the details of books available in the library. You can issue the books to the students and also maintain their records. Only one book is issued to one student. New books are not issued to students those not returned the last book. We can add, delete, search and modify the records of the file.

TABLE OF CONTENTS

SL No	Description	Pg No
1.	INTRODUCTION	
1.1	Introduction to File Structures	1
1.2	Introduction to Indexing	1
1.3	Scope and importance of work	2
1.4	Existing system and need for this Application	2
1.5	Tools used	2
2	IMPLEMENTATION	
2.1	Algorithm	3-4
2.2	Problems faced during implementation	4
3.	INDEXING ANALYSIS	
3.1	Analysis	5-6
4	RESULTS AND SNAPSHOTS	
4.1	Snapshots	7-14
4.2	Result	15
5	CONCLUSION AND REFERENCES	
5.1	Conclusion	16
5.2	References	16



CHAPTER-1 INTRODUCTION

1.1 Introduction to File Structures

File Structures is the Organization of data in secondary storage device in such a way that minimize the access time and the storage space. A File structure is a combination of representations for data in files and of operations for accessing the data. A File Structure allows applications to read, write and modify data. It might also support finding the data that matches some search criteria or reading through the data in some particular order.

- Since the details of the representation of the data and the implementation of the operations determine efficiency of the file structure for particular applications, improving these details can help improve secondary storage access time.
- Get the information we need with one access to the disk.
- If that's not possible, then get the information with as few accesses as possible.
- Group information so that we are likely to get everything we need with only one trip to the disk.

1.2 Introduction to Indexing

Indexing is the process of associating a key with the location of a corresponding data record. Index is a structure that contains a set of entries, each consisting of a key field and a reference field, which is used to locate records in a data file. Key field is the part of index that contains keys. Reference field is a part of the index that contains information to locate records. Primary keys are unique and are used to identify the records. Secondary key is the key other than the primary key and the index built on secondary key is secondary index. Simple indexing can be useful when the entire index can be held in memory. Changes require both the index and the data file to be changed. Updates affect the index if the key field is changed, or if the record is moved. An update which moves a record can be handled as deletion followed by an addition .

1.3 Scope and Importance of Work

The scope of the project is to access the record of students and books in minimum access to the disk. This project retrieves data, search, insert, delete and modify the records by taking less time thus reducing the access to the disk.

In this application we are able to save the records related to books and the students who have borrowed it.

This application is important. As the earlier method of accessing and saving the data records would take a lot of time and more access to the disk. With the help of indexing, the indexes are extracted and stored in main memory. So only the specific indexed records are retrieved.

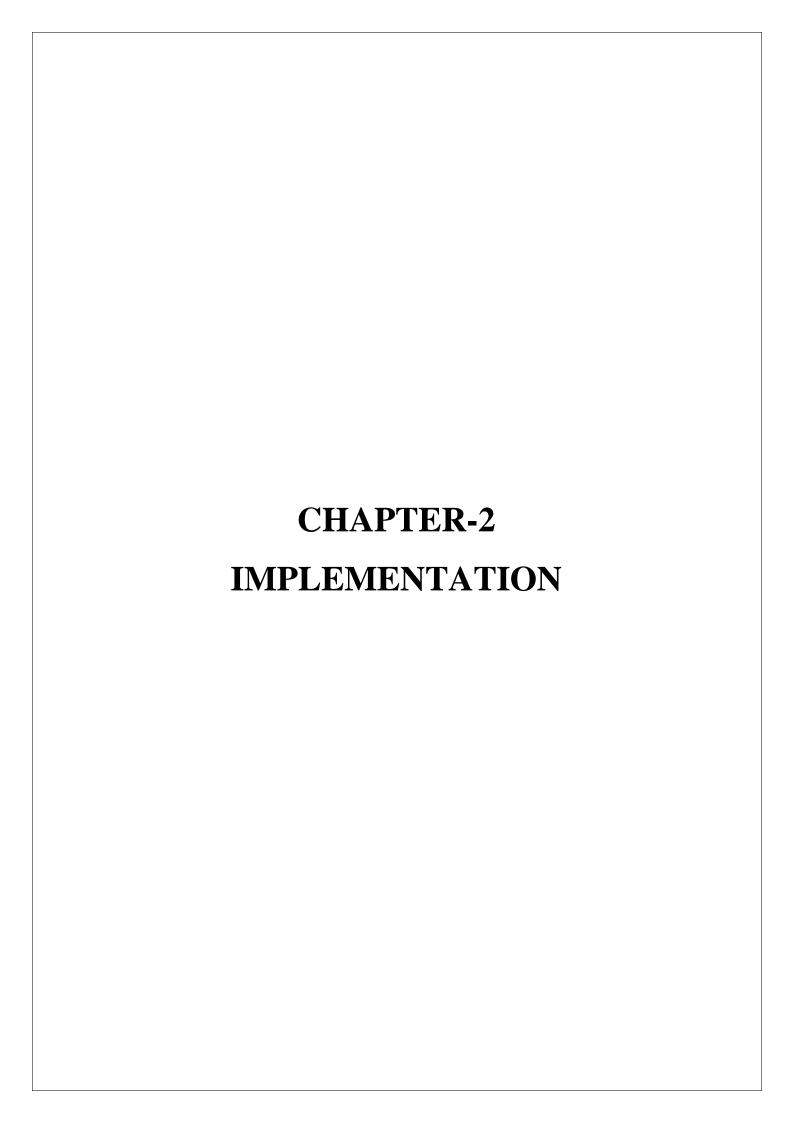
1.4 Existing application and need for this application

The existing application takes multiple accesses to the disk to retrieve and to perform other operations on the library dataset. This takes a lot of time. But this application minimizes the access to the disk. Time taken is less. The advantages of this application are:

- Easy to retrieve the data
- Easy to search
- Modification is simpler
- Minimum accesses to the disk

1.5 Tools Used

RAM:2.00 GB and more Java eclipse



CHAPTER-2 IMPLEMENTATION

2.1 Algorithm

- Step 1: We need to create a file containing a particular dataset
- Step 2: We need to take a search key as input
- Step 3: Output: Efficiently returns a collection of matching records.
- Step 4: The first column is the search key that contains the copy of the key index.
- Step 5: The second column is the pointer which contains a set of pointers holding the address of the disk block where that particular key value is found.

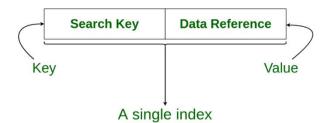


Fig 2.1:Parts of indexes. Search key and Reference part

- Step 6: Record addition: This consists of appending the data file and inserting a new record. The rearrangement of the index consists of sliding down the records with keys larger than thee inserted key and then placing new records in the opened space.
- Step 7: Record Deletion: This should use the techniques of reclaiming the space in file when deleting from the file. We must delete the corresponding entry from the index. Shift all records with key larger than key of the deleted record to the previous position in memory or make the index entry as deleted by using special character as *,- ,etc.
- Step 8: In my record file the primary index is book id and the secondary indexes are book name and author name.
- Step 9: Record addition in secondary indexing: When adding a record entry must also be added to the secondary key indexes. There may be duplicates in secondary key, keep duplicates in sorted order of primary key.

Indexing for Library data [Chapter-2]

Step 10: Record deletion in secondary indexing: Deleting a record implies removing all the references to the record in primary index and in all secondary indexes. When accessing the file through secondary key, the primary indexed file will be checked and a deleted record can be identified.

Step 11:Record Modification: Modifying for the data records is done by searching based on primary index. After that choice is given to modify the desired field. If the length of the new field is less than old field, the data is written in that position. If it exceeds then, the older data record is deleted or marked deleted and write the new record at the end of the file.

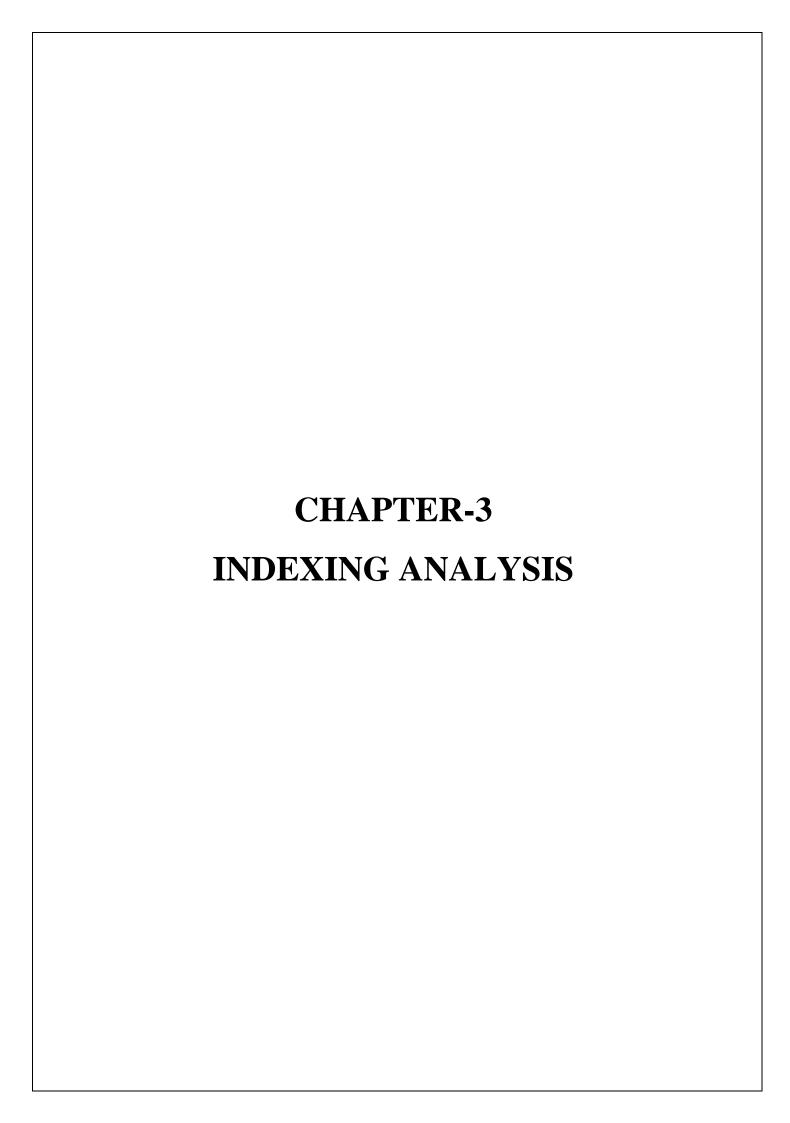
Step 12:Search: It allows binary search to obtain a keyed access to a record in variable length record file.

Step 13:Time taken for dataset has been calculated for each functionality.

2.2 Problems faced during the implementation

There were many problems faced by me during this implementation:

- Firstly, the csv file was not imported correctly and after imported the data was not retrieved properly.
- The next problem I faced was searching from 1 lakh data. There was an exception displayed while searching. There were many attempts that went in vain. Finally with the usage of trim(), this exception was solved and the search function worked.
- There were many small problems that were raised during implementation and those were solved at the moment itself.
- There was an exception displayed while deleting. That was raised due to space in the string and that string was converted to long which is not valid input. This exception was also solved using trim() function.



CHAPTER-3 INDEXING ANALYSIS

3.1 Analysis

Time efficiency for this application is showed by the below graph:

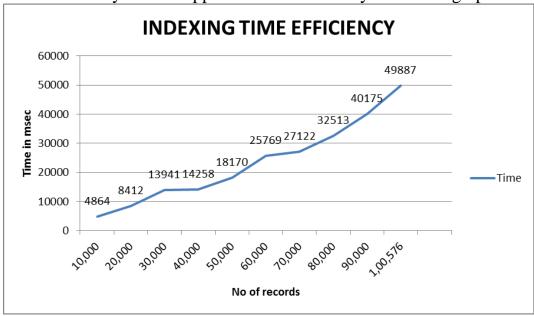


Fig 3.1 Time Efficiency graph of Indexing

From the above graph, it is observed that as we increase the number of records, the time in milli seconds were also increased. So by this we can say that the time required for indexing is directly proportional to the number of input records. Here are some cases that acts as an example for this explanation:

- For 10,000 records, the time taken to form the index is 486 ms.
- As we increase the number to 20,000 records, the time was also increased to 8412 ms.
- When number of records were 30,000, the time for indexing was 13941ms
- When number of records were 40,000, the time for indexing was 14258 ms.
- When number of records were 50,000, the time for indexing was 18170ms
- When number of records were 60,000, the time taken for indexing was 25769 ms.
- When the number of records were 70,000, the time taken for indexing was 27122 ms.
- When the number of records were 80,000, the time taken was 32513 ms.
- When the number of records were 90,000, the time taken was 40175 ms.

• When the number of records were approximately 1 lakh, the time taken was 49887 ms.

So analysis that can be drawn from the above observation is that number of records and time required for indexing are directly related.

This time efficiency graph can also be represented in the form of bar graph.

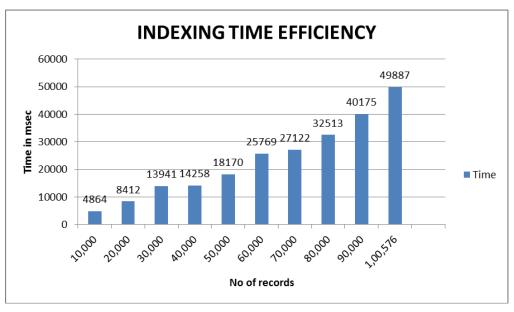
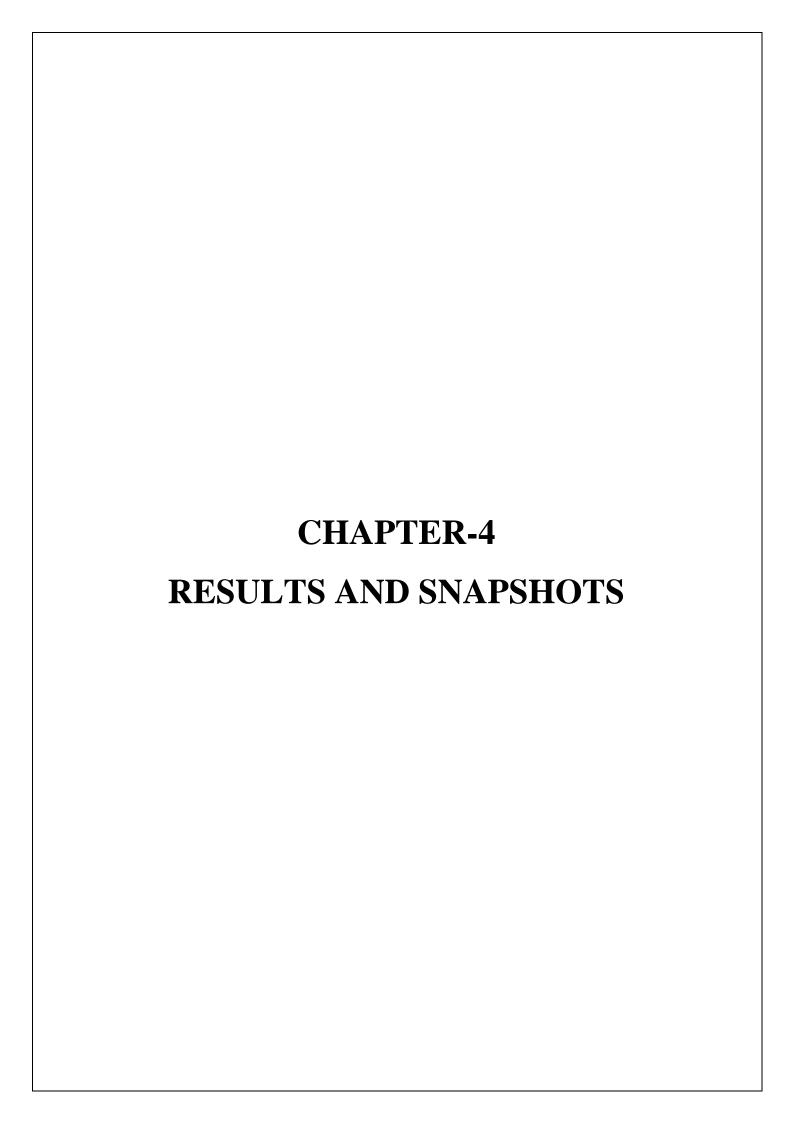


Fig 3.2:Bar graph representation of indexing time efficiency



CHAPTER-4 RESULTS AND SNAPSHOTS

4.1 Snapshots

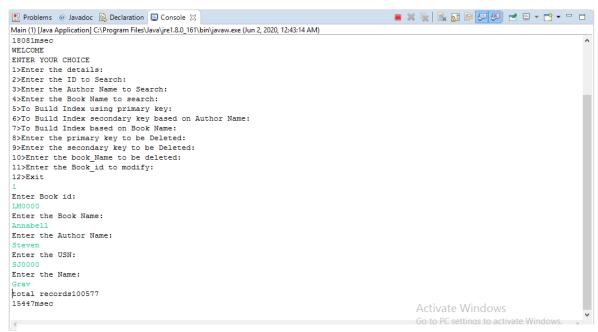


Fig 4.1:Insertion of record to the file: This window is displayed when the user enters the choice 1.

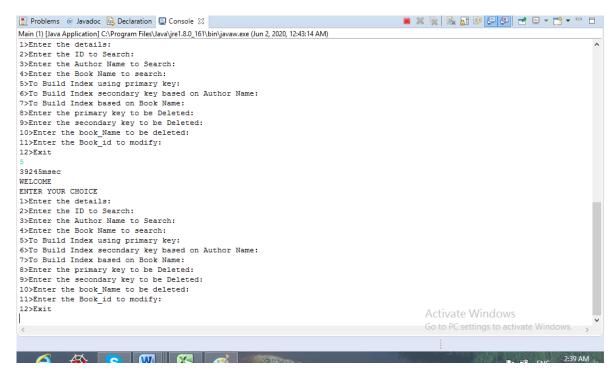


Fig 4.2: Updation of primary index: When the user enters 5 as the choice to update the primary index after inserting new record.

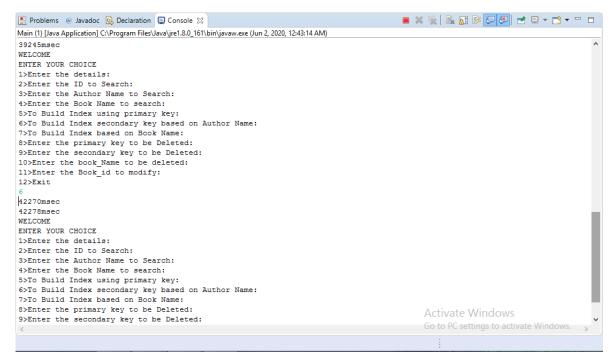


Fig 4.3:Updation of Secondary Index: This screen represents what happens when user enters 6 as the choice to update the secondary index which is author name

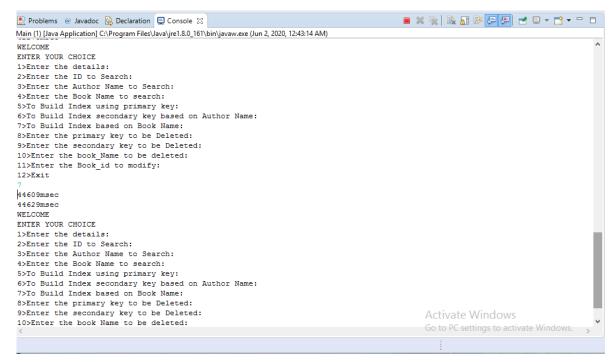


Fig 4.4:Updation of secondary index:This screen represents what happens when user enters 7 as the choice to update secondary index which is book name.

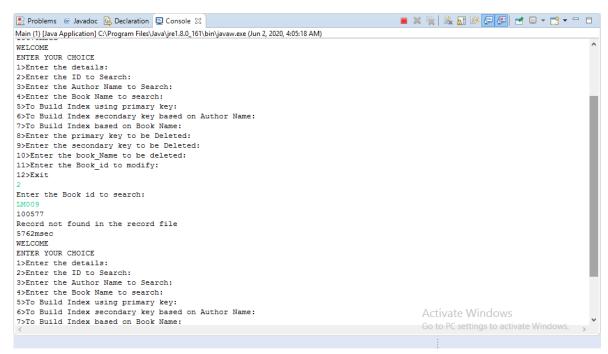


Fig 4.5:Search by Book_id:This screen represents what happens when user wants to search record that doesn't exist by choosing 2 as choice

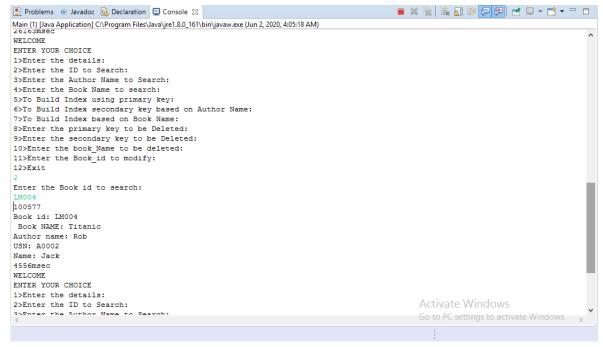


Fig 4.6:Searching using Book-id: This screen represents what happens when the user chooses 2 as choice and enters the book-id as LM004 whose details exist

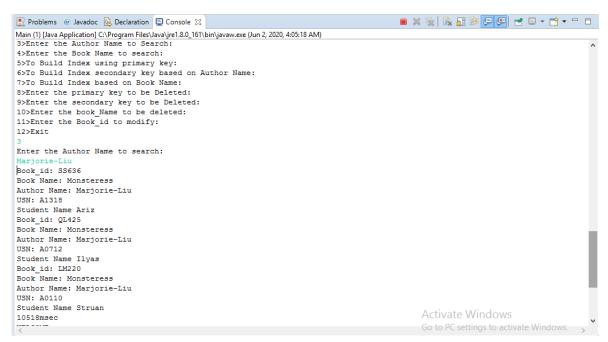


Fig 4.7:Searching using Author_Name:This screen represents what happens when user wants to search the records and chooses 3 as choice and also enters the author name which is Marjorie-Liu. It displays 3 records

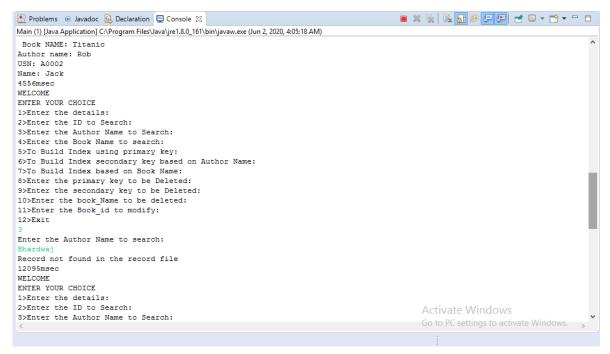


Fig 4.8:Searching using Author-Name:This screen represents what happens when user enters choice as 3 and enters Author Name as Bhardwaj but the record dosent exist

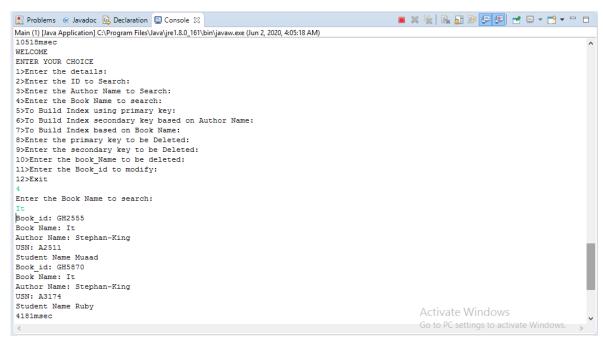


Fig 4.9:Searching based on book name: This screen represents what happens when the user enter choice 4 to search based on book name and enters book name as IT which results in two records.

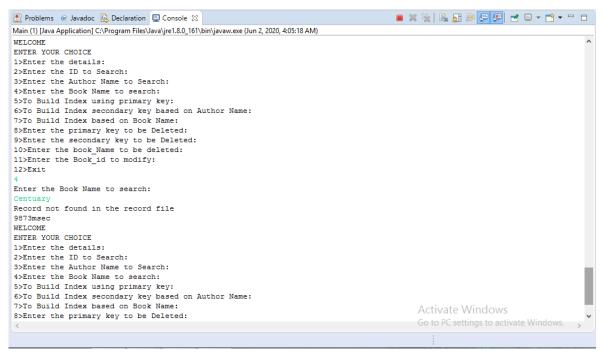


Fig 4.10:Searching using Book Name:This screen represents what happens when user enters choice 3 and book name as Centuary but that record doesn't exist.

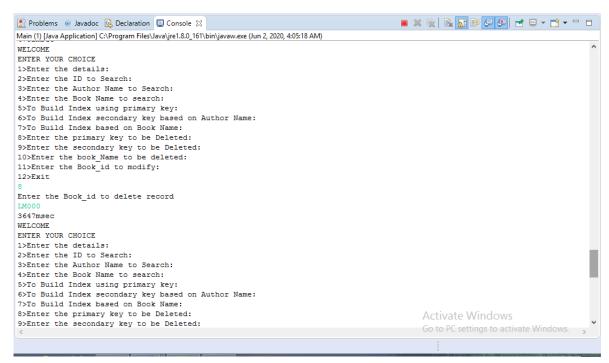


Fig 4.11:Deleting using primary key:This screen represents what happens when user enters choice as 8and also enters book id as LM000



Fig 4.12:Searching the deleted record:When the user enters the choice as 2 and enters the book id that is already deleted(LM000 in this case). It has been deleted is displayed.

Indexing for Library data[Chapter-4]

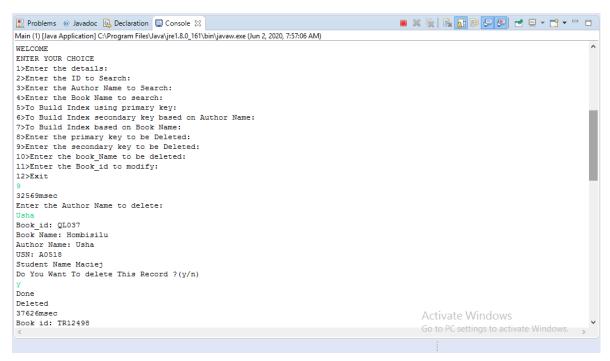


Fig 4.13:Deletion using Author Name:The user enters the choice as 9 to delete using author_name and enters author name as Usha. So the message displayed will be done deleted

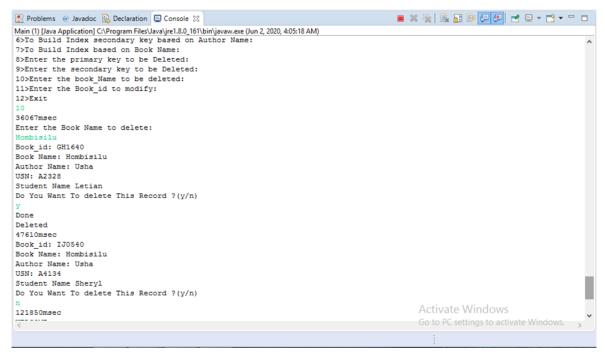


Fig 4.14Deletion using Book name: The user enters choice as 10 and enters Hombisilu as book name

Indexing for Library data[Chapter-4]

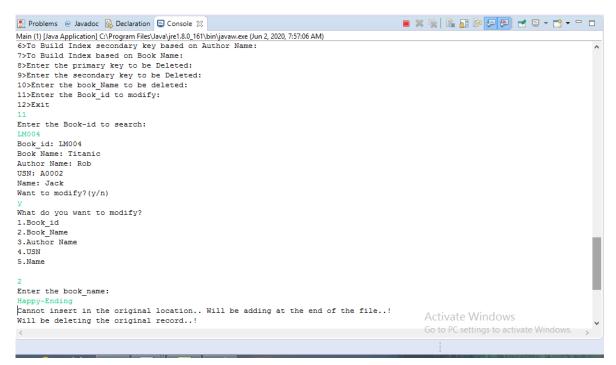


Fig 4.15:Modification of record: User enters choice as 11 and enters the book id to search. After entering book id the options of fields that has to be modified is displayed. The user selects book name field and enters the new name of the book. Since the length is more than old one, the old one is marked deleted and the new one is appended to the field

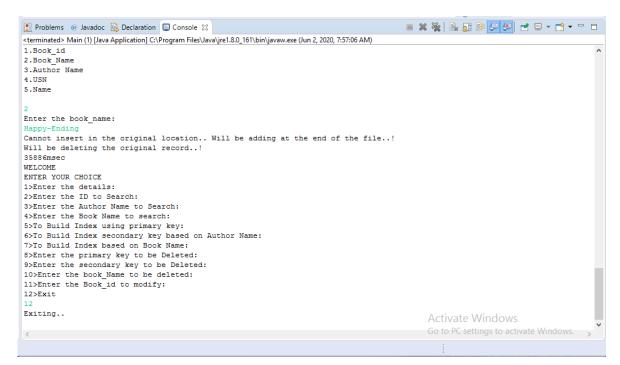


Fig 4.16:Exiting: When the user wants to quit from the project then, he chooses 12 as choice

Indexing for Library data[Chapter-4]

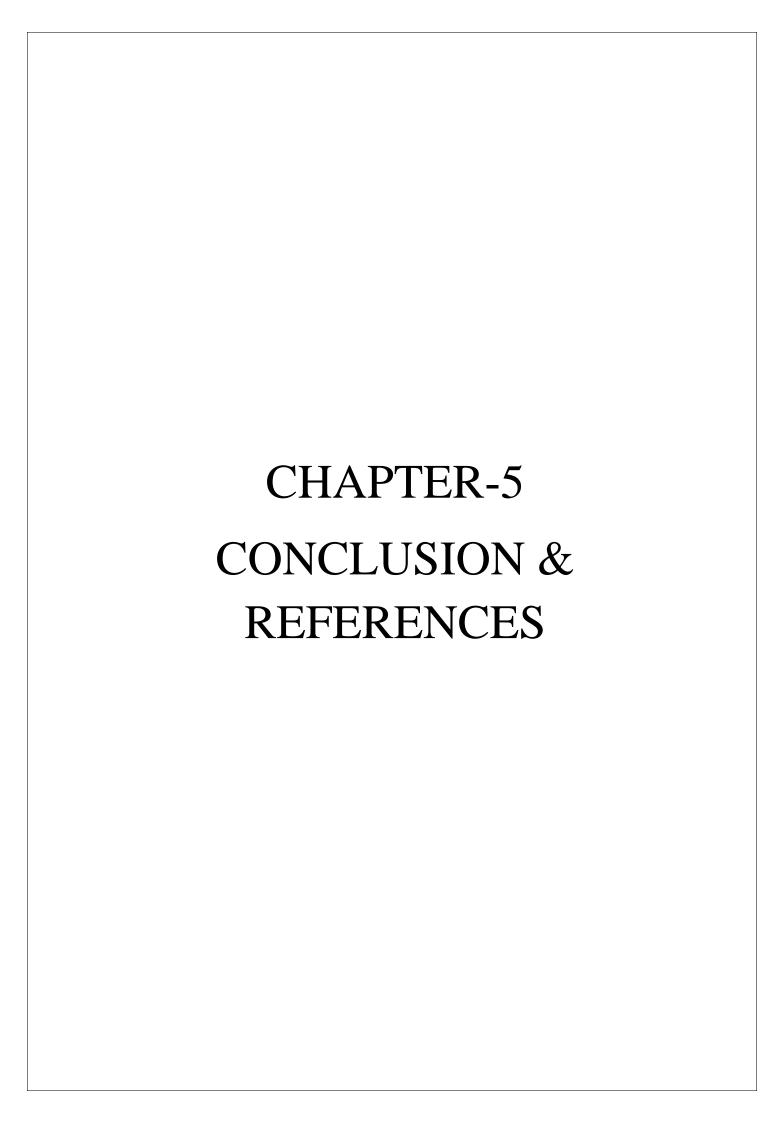
4.2 Result

This project solves the purpose for which it was designed. The operations that can be done using this project are:

- Insertion of new record to the existing file.
- Searching using both primary index(book_id) and secondary indexes(book_name & author_name)
- Deletion of the existing records using primary index and secondary indexes.
- Modifying the desired fields using book_id.

These can be achieved with few accesses to the disk and minimum time.

There are few observations in the form of snapshots attached



CHAPTER-5 CONCLUSION AND REFERENCES

5.1 Conclusion

We can maintain the record details of library. There is a scope for introducing a methods to insert a book which in turn can be developed as Student-Library system.

The purpose of the project was to build an application to reduce the manual work for managing library details. This application is used to store library details where in it holds some operations such as insertion, deletion, selection and modifications and so on. This application takes minimum access, any future enhancements can be done.

5.2 References

- 1. JAVA TUTORIAL-TUTORIALSPOINT: This site has helped me in learning some programming techniques in java. This helped me to implement few operations to the project like insertion, deletion, selection and modification.
- 2. JAVAPOINT: This site has helped me to learn of how to import csv file to eclipse.
- 3. GEEKSFORGEEKS: This site has helped me to understand and implement few theoretical concepts related to indexing to the project.