## Visvesvaraya Technological University Belagavi-590 018, Karnataka



A Mini Project Report on

## “Implementation of Indexing on e-books dataset”

## Mini Project Report submitted in partial fulfilment of the requirement for the File Structures Lab [17ISL68]

**Bachelor of Engineering In**

**Information Science and Engineering Submitted by**

**Arathi N [1JT17IS006] Under the Guidance of Mr.Vadiraja A**

**Asst. Professor**

**Dept. Of ISE**



**Department of Information Science and Engineering Jyothy Institute of Technology**

**Tataguni, Bengaluru-560082 2020-21**

**Jyothy Institute of Technology Tataguni, Bengaluru-560082**

**Department of Information Science and Engineering**



**CERTIFICATE**

Certified that the mini project work entitled **“Implementation of Indexing on e-books dataset”** carried out **by Arathi N [1JT17IS006]** bonfire 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 **2020-2021**. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the Report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of Project work prescribed for the said Degree.

### Mr .VadirajaA Dr. HarshvardhanTiwari

Guide,Asst.Professor Associate. Professor andHOD

Dept.Of ISE Dept.Of ISE

ExternalVivaExaminer Signature with Date: 1.

2.

## ACKNOWLEDGEMENT

Firstly, we are very grateful to this esteemed institution **“Jyothy Institute of Technology**” for providing us an opportunity to complete our project.

We express our sincere thanks to our Principal **Dr. Gopalakrishna K for** providing us with adequate facilities to undertake this project.

We would like to thank **Dr. Harshvardhan Tiwari , Associate Prof. and Head** of Information Science and Engineering Department for providing for his valuable support.

We would like to thank our guides **Mr.Vadiraja A** , **Asst. Prof.** for their keen interest and guidance in preparing this work.

Finally, we would thank all our friends who have helped us directly or indirectly in this project.

### Arathi N[1JT17IS006 ]

**ABSTRACT**

Indexing is the process of associating a key with the location of a corresponding data record.

An external sort typically uses the concept of a key sort,in which an index file is created wh-

ose records consist of keypairs.Here,each key is associated with a pointer to a complete rec-

ord in the main database file.The index file could be sorted or organised using a tree structure,

thereby imposing a logical order on the records without physically rearranging them.Each rec-

ord of a database normally has a unique identifier ,called the primary key.A particular key v-

alue might be duplicated in multiple records,is called a secondary key.The secondary key ind-

ex will associate a secondary key value with the primary key of each record having that seco-

ndary key value .The full database might be searched directly for the record with that primary

key,or there might be a primary key index that relates each primary key value with a pointer

to the actual record on the disk.In this case,the primary index provides the location of the act

-ual record on disk,while the secondary disk indices refer to the primary index.Indexing is an

important technique for organising large databases.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **SL.NO** | **DESCRIPTION** | **PG NO.** |
|  | **Chapter 1** |  |
| 1. | **Introduction** | 1 |
| 1.1 | Introduction to File Structure | 2 |
| 1.2 | Introduction to Java | 2 |
| 1.3 | Introduction to Indexing | 3 |
|  |  |  |
|  | **Chapter 2** |  |
| 2. | **Implementation** | 5 |
| 2.1 | Basic operations on Indexing | 6 |
| 2.2 | Algorithm | 6 |
|  |  |  |
|  | **Chapter 3** |  |
| 3. | **Indexing Analysis** | 8 |
| 3.1 | Indexing Analysis | 9 |
|  |  |  |
|  | **Chapter 4** |  |
| 4. | **Results and Snapshots** | 10 |
| 4.1 | Primary Indexing | 11 |
| 4.2 | Entering the details | 11 |
| 4.3 | Searching based on primary key | 12 |
| 4.4 | Index Building | 12 |
| 4.5 | Deletion based on primary key | 13 |
| 4.6 | Exit | 13 |
| 4.7 | Secondary Indexing | 14 |
| 4.8 | Entering the details | 14 |
| 4.9 | Searching based on secondary key | 15 |
| 4.10 | Index Building | 15 |
| 4.11 | Deletion based on secondary key | 16 |
| 4.12 | Exit | 16 |
|  |  |  |
|  | **Conclusions** | 17 |
|  | **References** | 18 |

CHAPTER 1

INTRODUCTION

**1. INTRODUCTION**

IMPLEMENTATION OF INDEXING

**1.1 Introduction to File Structure**

* A disk’s relatively slow access time and the enormous, nonvolatile capacity is the driving force behind FILE STRUCTURE design!!
* FS should give access to all the capacity without making the application spend a lot of time waiting for the disk.
* FS is a combination of representation for data in files and of operations for accessing the data.
* It allows applications to read, write and modify data
* Also finding the data
* Or reading the data in a particular order
* Efficiency of FS design for a particular application is decided on,
* Details of the representation of the data
* Implementation of the operations
* A large variety in the types of data and in the needs of application makes FS design important.
* What is best for one situation may be terrible for other.

**1.2 Introduction to Java**

* **Java** is a general-purpose computer-programming language that is concurrent, class-based, object oriented and specifically designed to have as few implementation dependencies as possible. It is intended to let application developers "write once, run anywhere" (WORA), meaning that compiled Java code can run on all platforms that support Java without the need for recompilation.

.

* Java applications are typically compiled to byte code that can run on any Java virtual machine (JVM) regardless of computer architecture.

DEPT.OF ISE,JIT

2

2020-21

* James Gosling, Mike Sheridan, and Patrick Noughton initiated the Java language project in June 1991. Java was originally designed for interactive television, but it was too advanced for the digital cable television industry at the time.The language was initially called *Oak* after an oak tree that stood outside Gosling's office. Later the project went by the name *Green* and was finally renamed *Java*, from Java Coffee. Gosling designed Java with a C/C++-style syntax that system and application programmers would find familiar.

IMPLEMENTATION OF INDEXING

Sun Microsystems released the first public implementation as Java 1.0 in 1996. It promised "Write Once, Run Anywhere" (WORA), providing no-cost run-times on popular platforms.

* There were five primary goals in the creation of the Java language:
* It must be "simple, object-oriented, and familiar".
* It must be "robust and secure".
* It must be "architecture-neutral and portable".
* It must execute with "high performance".
* It must be "interpreted, threaded, and dynamic"

**1.3 Introduction to Indexing**

* Indexing is a data structure technique which allows you to quickly retrieve records from a database file.
* An Index is a small table having only two columns. The first column comprises a copy of the primary or candidate key of a table. Its second column contains a set of pointers for holding the address of the disk block where that specific key value stored.
* An index takes a search key as input and efficiently returns a collection of matching records.

Type of Indexes:

* Primary Indexing - Primary Index is an ordered file which is fixed length size with two fields. The first field is the same a primary key and second, filed is pointed to that specific data block. In the primary Index, there is always one to one relationship between the entries in the index table.

3

2020-21

DEPT OF ISE,JIT

* Secondary Indexing -The secondary Index can be generated by a field which has a unique value for each record, and it should be a candidate key. It is also known as a non-clustering index.

IMPLEMENTATION OF INDEXING

4

2020-21

DEPT OF ISE,JIT

4

3

2020-21

DEPT OF ISE,JIT

CHAPTER 2

IMPLEMENTATION

**2.1 Basic operations on Indexing**

IMPLEMENTATION OF INDEXING

In this section, we present the details of the operations of indexing:

* Entering the details.
* Searching.
* Deleting.
* Build Index

**2.2 Algorithm**

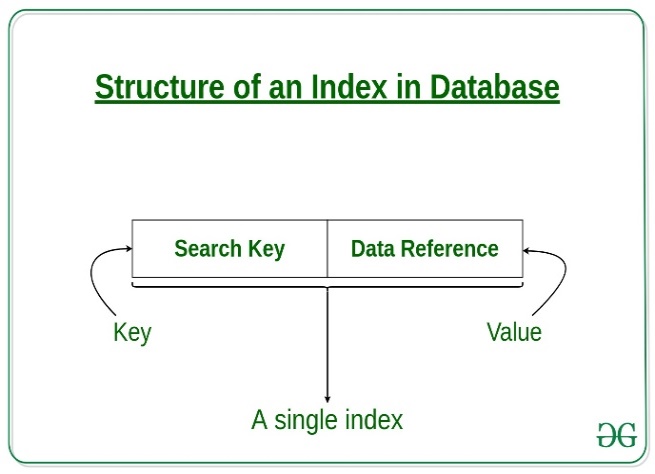
Step : Creation of a file containing a particular dataset.

Step 2: Input -Takes 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 a copy of primary key or candidate key of the table.

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** 2.2.1

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 the inserted key and then placing new record in the opened space.

Step 7: Record deletion-This should use the techniques for reclaiming space in files when deleting from the data file. We must delete the corresponding entry from the index . Shift all records with keys larger than key of the deleted record to the previous position in memory or make the index entry as deleted.

6

2020-21

DEPT OF ISE,JIT

Step 8: In our record file we built an index for b\_id which is primary key and there is author name as secondary key.

IMPLEMENTATION OF INDEXING

Step 9: Record addition in secondary indexing :When adding a record entry must also be added to the secondary key index. There may be duplicates in secondary key, keep duplicates in sorted order of primary key.

Step 10: Record deletion in secondary key : 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: It allows binary search to obtain a keyed access to a record in variable length record file.

Step 12: Time taken for dataset has been calculated for each functionality.

7

2020-21

DEPT OF ISE,JIT

7

2020-21

DEPT OF ISE,JIT

CHAPTER 3

INDEXING ANALYSIS

IMPLEMENTATION OF INDEXING

**3.1 Indexing Analysis**

The graph mentioned below gives the time analysis for no of records versus time taken for

Index building in ms.In the first case,for 10,000 record set it takes time around 30,000ms for index building.In the second case,for 20,000 record set it takes time around 50,000ms for index building.In the third case,for 30,000 record set it takes time nearing to 1,00,000ms for index building.In the fourth case,for 50,000 record set it takes time nearing to 1,50,000ms for index building. In the fifth case,for 10,00,000 record set it takes time nearing to 3,00,000ms for index building.

**Fig** 3.3.1 Time analysis for indexing

2020-21

DEPT OF ISE,JIT

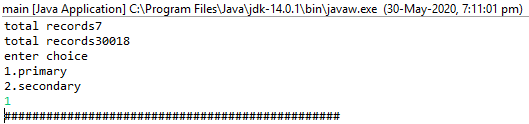
9

CHAPTER 4

RESULTS AND SCREENSHOTS

**4.1 Primary Indexing**

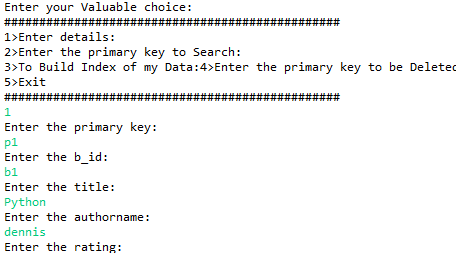
IMPLEMENTATION OF INDEXING



**Fig** 4.4.1 Primary Indexing

By selecting option 1 you can perform primary indexing and by selecting option 2 you can perform secondary indexing.

**4.2 Entering the details**

****

**Fig** 4.4.2 Entering the details

In the above image the user has selected primary indexing and further has selected choice 1 in which you can enter details of e-books dataset.The time taken to enter all the details here is 101532ms.

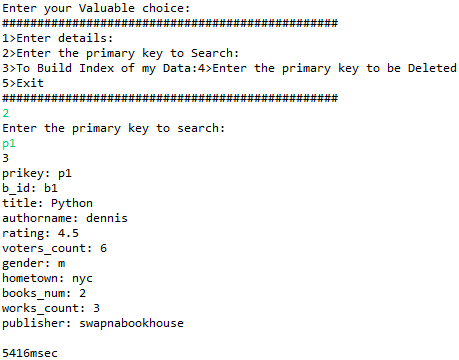
11

2020-21

DEPT OF ISE,JIT

**4.3 Searching based on primary key**

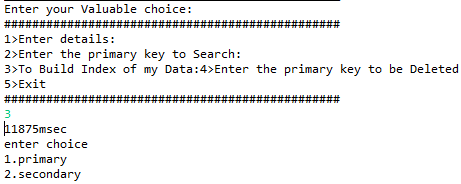
IMPLEMENTATION OF INDEXING

****

**Fig** 4.4.3 Searching based on primary key

Here the user has selected choice 2 in which you can search based on the primary key.

**4.4 Index Building**

****

**Fig** 4.4.4 Index Building

The user has selected choice 3 in which you can build the index of the data which takes 11875msec.

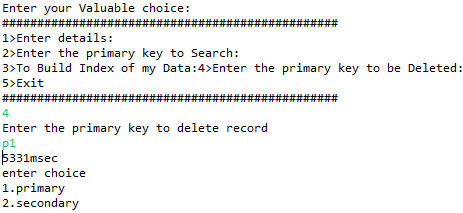
DEPT OF ISE,JIT

12

2020-21

**4.5 Deletion based on primary key**

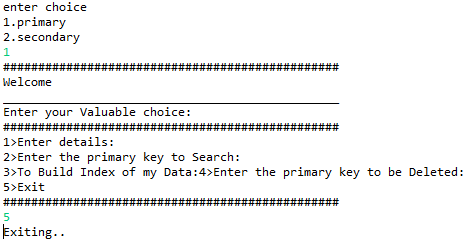
IMPLEMENTATION OF INDEXING

****

**Fig** 4.4.5 Deletion based on primary key

In this image user has selected choice 4 in which you can perform deletion using primary key(b\_id).

**4.6 Exit**

****

**Fig** 4**.**4.6 Exit

Here user selects choice 5 and exits.

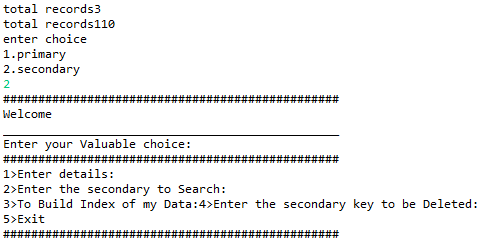
13

2020-21

DEPT OF ISE,JIT

**4.7 Secondary Indexing**

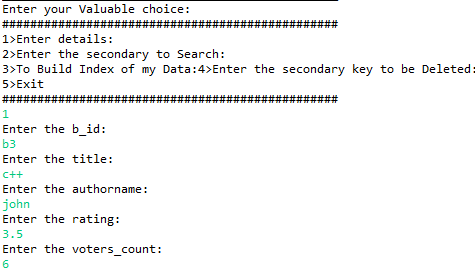
IMPLEMENTATION OF INDEXING

****

**Fig** 4.4.7 Secondary Index

In this image user has selected choice 2 that is secondary indexing.

**4.8 Entering the details**

****

**Fig** 4.4.8 Entering the details

Here user has selected choice 1 in which you can enter the details of the dataset.

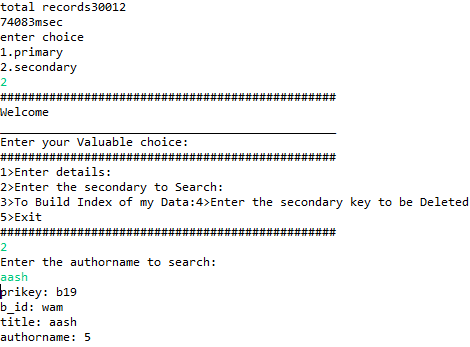
14

2020-21

DEPT OF ISE,JIT

**4.9 Searching based on secondary key**

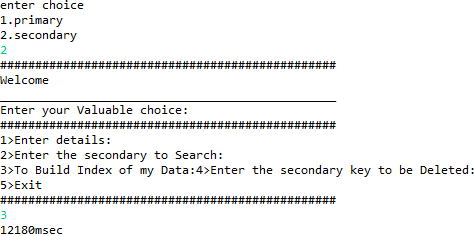
IMPLEMENTATION OF INDEXING

****

**Fig** 4.4.9 Searching based on secondary key

In this image user has selected choice 2 in which searching of the data can be done based on secondary key(authorname).

**4.10 Index Building**

****

**Fig** 4.4.10 Index Building

Here building of the index is done which takes 12180msec.

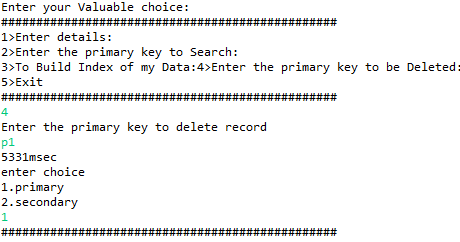
15

2020-21

DEPT OF ISE,JIT

**4.11 Deletion based on secondary key**

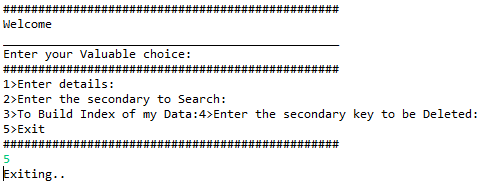
IMPLEMENTATION OF INDEXING

****

**Fig** 4.4.11 Deletion based on secondary key

In this image user has selected choice 4 in which deletion operation based on the secondary key is done.

**4.12 Exit**

****

**Fig** 4**.**4.12 Exit

Here user selects choice 5 and exits.

**CONCLUSION**

DEPT OF ISE,JIT

2020-21

16

We have successfully implemented indexing which helps us in administrating the data used for managing the tasks performed.

IMPLEMENTAION OF INDEXING

View tables are used to display all the components at once so that user can see all the components of a particular type at once. One can just select the component and modify and remove the component.

We have successfully used various functionalities of JAVA and created the File structures.

Features:

1. Clean separation of various components to facilitate easy modification and revision.

2. All the data is maintained in a separate file to facilitate easy modification

3. All the data required for different operations is kept in a separate file.

4. Quick and easy saving and loading of database file.

DEPT OF ISE,JIT

2020-21

17

**REFERENCES**

IMPLEMENTATION OF INDEXING

The information about Indexing was gathered by referring to the following sites**:**

* Github(github.com)
* Wikipedia(www.wikipedia.org)
* Stackoverflow(stackoverflow.com)
* We3schools(Wethreeschools.com)
* GeeksforGeeks(GeeksforGeeks.com)
* Javatpoint(javatpoint.com)
* Java Debugger(javadebugger.com)
* Youtube(www.youtube.com)
* Michael J.Folk,Bill Zoeclick,Greg Riccardi:File Structures-An Object Oriented Approach with C++,3rd Edition,Pearson Education ,1998.
* Scott Robert Ladd:C++ Components and Algorithms,BPB Publications,1993.

18

2020-21

DEP OF ISE,JIT