Progressive Education Society's
Modern College of Engineering
DEPARTMENT OF COMPUTER ENGINEERING

ONE DAY ONLINE SYLLABUS IMPLEMENTATION

FACULTY DEVELOPMENT PROGRAM

ON

DATA SRTUCTURES AND ALGORITHMS

SE COMP, 2019 Course

Date – 18th January 2021

Ms. Pallavi V. Baviskar

Assistant Professor, Computer Engineering Department

PES MCOE Pune-5

Unit VI: File Organization

Mid-Sem :30
End-Sem :70

Hours: 07
3 lectures\week

Syllabus:

Files: concept, need, primitive operations.

Sequential file organization- concept and primitive operations

Direct Access File- Concepts and Primitive operations

Indexed sequential file organization-concept, types of indices, structure of index sequential file,

Linked Organization- multi list files, coral rings, Inverted files and cellular partitions.
Content Beyond Syllabus: Random File Organization (Direct accessing, directory lookup, Hashing)

Case Study:

External Sort- Consequential processing and merging two lists, multiway merging- a k way merge algorithm

- **Objectives:**

To suggest appropriate file organization as solutions of the storage problems.

- **Outcome**

Explain different file organization.

Demonstrate merging two lists, multiday merging, k way merge algorithm

- **Mapping of Course outcomes for unit VI:**

Analyse the algorithmic solutions for resource requirement and optimisation.

Use appropriate modern tools to understand and analyse the functionalise confine to the secondary storage.

File organization

- Sequential organization
- Direct sequential organization
- Index sequential organization
- Linked organization
- Inverted files
- Cellular partitions

Record	E#	Name	Occup	Degree	Sx	Location	MS	Salary
A	800	Hawkins	Prgrmr	B.S.	M	Los Angeles	S	10000
B	510	Williams	Analyst	B.S	F	Los Angeles	M	15000
C	950	Frawley	Analyst	M.S.	F	Minneapolis	S	12000
D	750	Austin	Progrmr	B.S.	F	Los Angeles	S	12000
E	620	Messer	Progrmr	B.S.	M	Minneapolis	M	9000

Sample Employee File

Applications of File Organization

- Simple storage, surveillance, and retrieval of documents irrespective of location
- Superior search abilities mean less time is spent on searching for important files
- Lowered storage costs associated with physical files.
- More security and privacy control over sensitive company data
- Disaster recovery and backups
- Real-time collaboration and improved teamwork
- The version history of each document to see when edits were made
- File Organization -https://www.youtube.com/watch?v=GCO1erf9BLE

Sequential file organization

- https://www.youtube.com/watch?v=5czzGLlLjVY&feature=emb_logo

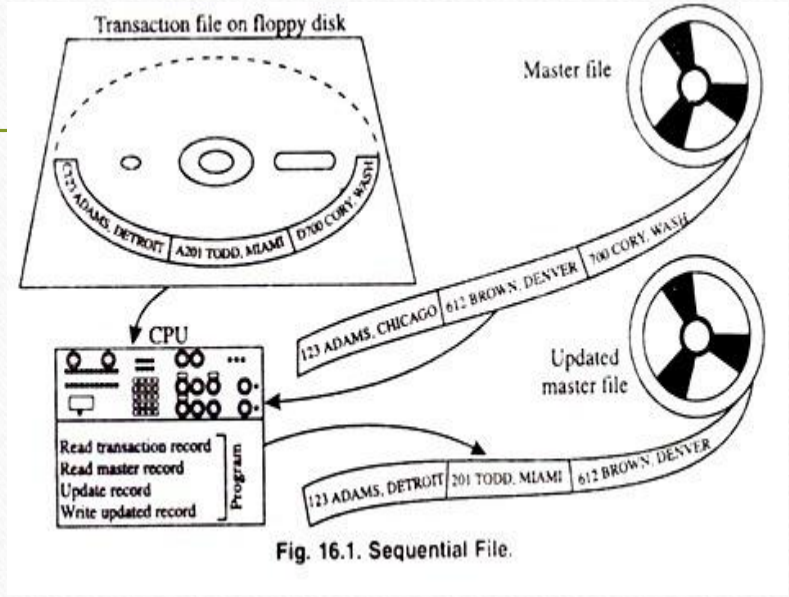


Fig. 16.1. Sequential File.

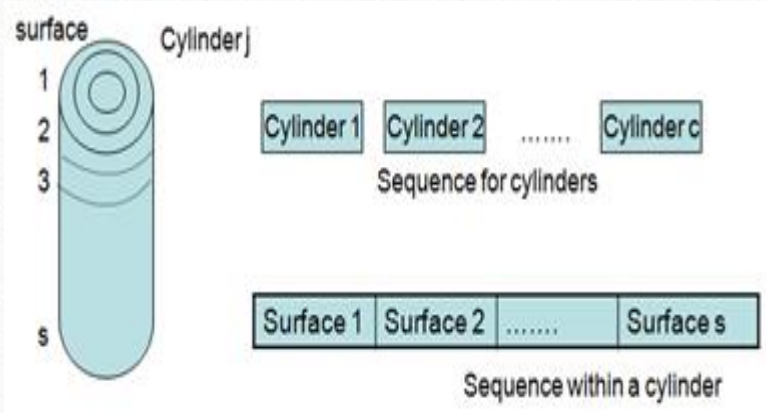
Advantages:

- Errors in files remain localized.
- Loading a record requires only the record key.
- Simple to understand approach.
- Easy to organize, maintain and understand.
- Relatively inexpensive input/output (I/O) media and device may be used.
- Files may be relatively easy to reconstruct since a good measure of built in back up is usually available.
- Efficient and economical if the activity rate, i.e. the proportion of file records to be processed is high.

Sequential file organization

Disadvantages:

- Transactions must be sorted and placed in sequence prior to processing.
- 'Data' redundancy is typically high since the same data may be stored in several files sequenced in different keys.
- Random enquiries are virtually impossible to handle.
- Timeliness of data in file deteriorates while batches are being accumulated.
- Entire file must be processed even when the activity rate is very low.
- This method is suitable for the small records. If the record is very big, so much time is consumed in arranging the file.



Direct sequential-File Organization

- https://www.youtube.com/watch?v=27KMiQfrATE

Advantages:

- Transactions need not be sorted.
- Different discs or disc units are not required for updating records as existing records may be amended by overwriting.
- It is also possible to process direct file records sequentially in a record key sequence.
- A direct file organization is most suitable for interactive on line applications such as air line or railway reservation systems, teller facility in banking application, etc.
- Immediate access to records for updating purposes is possible.
- Random inquiries which are too frequent in business situations can be easily handled.

Direct sequential-File Organization

Drawbacks:

- May be less efficient in the use of storage space than sequentially organized file.
- Expensive hardware and software resources are required.
- Relative complexity of programming.
- System design around it is complex and costly.
- Data may be accidentally erased or even written unless special precautions are taken.
- Special security measures are necessary for on line direct files that are accessible from several stations.
- File updating (addition and deletion records) is more difficult as compared to sequential files.
- Records in the on line may be expressed to the risks of loss of accuracy and a breach of security Special back up and reconstruction procedures must be established.

Index sequential-File Organization

Advantages of Indexed sequential access file organization:

- In indexed sequential access file, sequential file and random file access is possible.
- It accesses the records very fast if the index table is properly organized.
- The records can be inserted in the middle of the file.
- It provides quick access for sequential and direct processing.
- It reduces the degree of the sequential search.

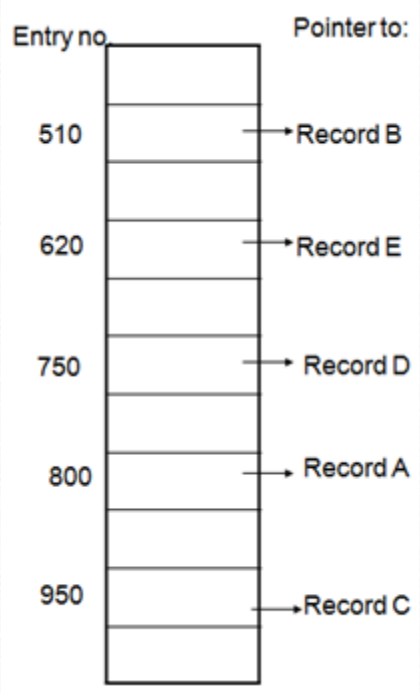
Index sequential-File Organization

Disadvantages of Indexed sequential access file organization

- Indexed sequential access file requires unique keys and periodic reorganization.
- Indexed sequential access file takes longer time to search the index for the data access or retrieval.
- It requires more storage space.
- It is expensive because it requires special software.
- It is less efficient in the use of storage space as compared to other file organizations

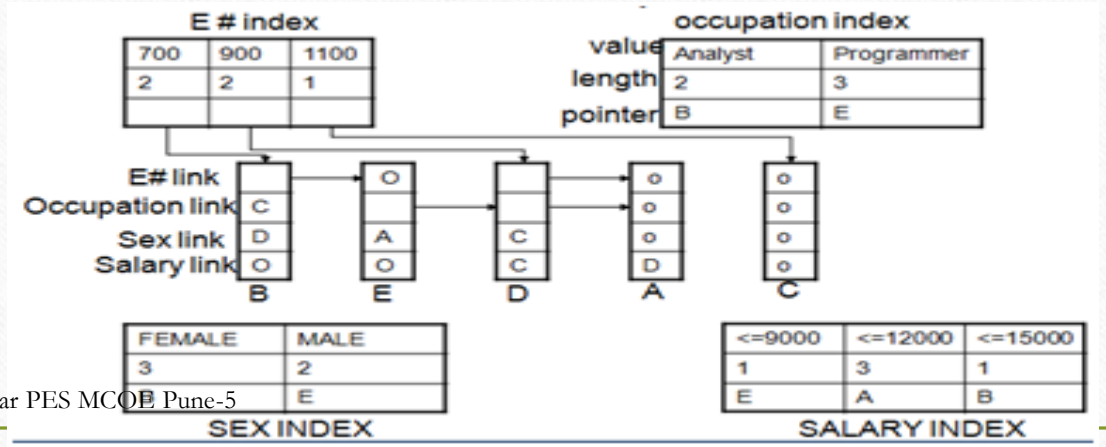
Random File Organisation

- Direct Accessing
- Directory lookup
- Hashing



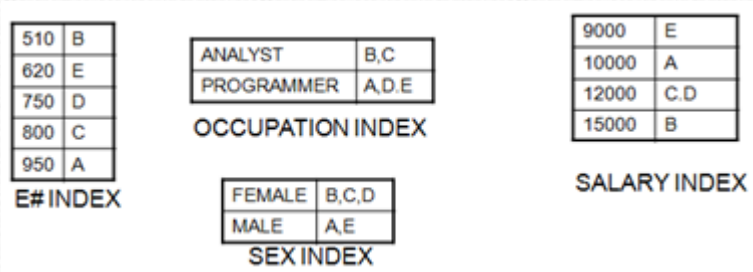
Linked Organization

- Linked organizations differ from sequential organizations essentially in that the logical sequence of records is generally different from the physical sequence.
- In sequential ith record is placed at location li, then the i+1st record is placed at li + c where c is the length of ith record or some fixed constant.
- In linked organization the next logical record is obtained by following link value from present record. Linking in order of increasing primary key eases insertion deletion.
- Searching for a particular record is difficult since no index is available, so only sequential search possible.



Inverted files

- Inverted files are similar to multilists. Multilists records with the same key value are linked together with link information being kept in individual record. In case of inverted files the link information is kept in index itself.
- The retrieval works in two steps. In the first step, the indexes are processed to obtain a list of records satisfying the query and in the second, these records are retrieved using the list.
- The no. of disk accesses needed is equal to the no. of records being retrieved + the no. to process the indexes.



Multi-list File Organization

- A multi-list is really only a slightly modified inverted file.
- There is one list per keyword, i.e. hi = 1. The records containing a particular keyword Ki are chained together to form the Ki-list and the start of the Ki-list is given in the directory.
- The multi-list is designed to overcome the difficulties of updating an inverted file.
- No such problem arises with the multi-list, we update the appropriate K-lists by simply chaining in the new record.
- The penalty we pay for this is of course the increase in search time.
- This is in fact typical of many of the file structures. Inherent in their design is a trade-off between search time and update time.

Multi-list File Organization

- Multi-list Vs Index file
- In a Multi-list organization an index entry points to the first data record in the list, whereas in inverted index file an index entry has address pointers to all the data records related to it.
- A multi-list index has fixed length records, whereas an inverted index contains variable length records

Cellular partitions

- To reduce the file search times, the storage media may be divided into cells.
- A cell may be an entire disk pack or it may simply be a cylinder. Lists are localized to lie within a cell.
- Thus if we had a multilists organization in which the list for key1=prog list included records on several different cylinders then we could break the list into several smaller lists where each prog list included only those records in the same cylinder.
- The index entry for prog will now contain several entries of the type (addr, length) where addr is a pointer to start of a list of records with key1=prog and length is the no. of records on the list. By doing this all records of the same cell may be accessed without moving the read/write heads.

Sample of University Questions

Solve any one question from Q.7(a,b) &Q.8(a,b)

7(a) Explain various file opening modes with respect to text and binary files.	6 M
7(b) What are the primitive operations on sequential file? Explain with example.	6 M
8(a) Compare the feature of sequential file, index sequential file and direct access file.	6 M
8(b) Write C++ perform the following operations on direct access file: i) Create & display records ii) Insert record.	6 M

Sample of MCQ's

- . _____ is a unique tag, usually a number identifies the file within the file system. **a) File identifier** b) File name c) File type d) None of the mentioned
- Which one of the following explains the sequential file access method? a) random access according to the given byte number **b) read bytes one at a time, in order** c) read/write sequentially by record d) read/write randomly by record
- When will file system fragmentation occur? **a) unused space or single file are not contiguous** b) used space is not contiguous c) unused space is non-contiguous d) multiple files are non-contiguous
- Airline reservation systems and inventory control system are the examples of …….. system. A sequential file B indexed sequential file **C indexed file** D none of the mentioned
- The ……. greatly reduced the time required to access a single record, without sacrificing the sequential nature of the file. A **indexed sequential file** B sequential file C pile D None of the above
- …… are typically used in batch applications and are generally optimum for such applications if they involve the processing of all the records. **A Sequential files** B Indexed Sequential files C Indexed files D Direct files
- ……. are often used where very rapid access is required, where fixed length records are used, and where records are always accessed one at a time. **A Direct files** B Indexed Sequential files C Indexed files D Sequential files

Learning Material

MOOC/ Video Lectures available at:

https://nptel.ac.in/courses/106/102/106102064/

https://nptel.ac.in/courses/106/105/106105085

https:// nptel.ac.in/courses/106/106/106106127

e-Books:

https://www.ebooks.com/en-us/book/95777110/Python-data-structures-and-algorithms/benjamin-baka/

https://www.ebookphp.com/advanced-data-structures-epub-pdf/

https://www.ebookphp.com/data-structures-and-algorithms-professional-edition-beginners-guide-epub-pdf/

References

Links:

- https://www.tutorialride.com/data-structures/file-organization-in-data-structure.htm
- https://www.tutorialspoint.com/data_structures_algorithms/index.htm

Books:

- Dr. Dobbs/books/book5/chap03.htm
- Aho, J. Hopcroft, J. Ulman, “Data Structures and Algorithms”
- Sartaj Sahani, “Data Structures, Algorithms and Applications in C++”

Papers:

- https://digital.library.unt.edu/ark:/67531/metadc663738/m2/1/high_res_d/1002772864-Rogers.pdf

Thank You…

Presented by, Ms. Pallavi V. Baviskar PES MCOE Pune-5