IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, January 2022
Impact Factor: 6.252

Analysis for Information on the Applicability of Data Structures

Suparna Roy[1] and Krishna Singh[2]
Assistant Professor, BSC CS, Suman Education Society's LN College, Borivali East, Mumbai, India[1]
Student, BSC CS, Suman Education Society's LN College, Borivali East, Mumbai, India[2]

Abstract: *This article aims to provide some insight on the application of data structures in the area of information retrieval. Given the increasing desire and necessity for sharing and examining knowledge, information retrieval is a field of research that is gaining ground. For a very long time, data structures have been the focus of research. epoch in the computer science field. Given the exponential growth of data, it is even more crucial to have effective data structures.*

Keywords: Data structures, Information retrieval

I. INTRODUCTION

Data has always been and continues to be a resource that should be used and used wisely for the benefit of businesses and institutions. With the rise of social networking sites and technical advancements, there is a huge amount of data sharing today. Information retrieval is the process of locating processed data from an existing repository. Information is defined as processed data. Multiple computer science disciplines are actively involved in the investigation of information retrieval. Information retrieval is a technology utilised in many advanced computer science study fields and makes use of many of the fundamental ideas in computer science. As an illustration, information retrieval is a preliminary stage in the text mining process before subsequent mining operations are used.

II. RETRIEVING INFORMATION

2.1 Information Extraction and Information Retrieval

The words information extraction (IE) and information retrieval (IR) are sometimes used interchangeably. They are utterly distinct domains with various, distinct jobs as the end result. Information extraction has no specific objectives or targets that must be met. It does make use of templates to give structure to otherwise unstructured data. Information retrieval calls for advanced methods since it must fulfil the user's demand to locate specific information from an existing repository. Selecting an appropriate index for better querying is another auxiliary function of information retrieval, and an intelligent information retrieval system employs user feedback to improve upon the current system and fine-tune the procedures. Data mining and information retrieval both use the same summarization and clustering techniques.

2.2 The Effectiveness of an IR System

Based on the system's reaction time and quality, an information retrieval system's success rate or performance is determined. of the result. Another qualitative word that may be assessed by user input is the quality of the information retrieval response. Precision and recall are the typical measures employed for the quality measurement. The percentage of retrieved relevant documents to the total number of relevant documents is the definition of the recall metric. The proportion of pertinent papers found among all documents found constitutes the precision metric. The relevance of the materials is a qualitative phrase even if these measurements have clear meanings.

A quantitative metric is the information retrieval system's reaction time since it can be measured. The elements that have an impact on the reaction time of the size and organisation of the corpus to be searched, the kind of index being utilised, and the kind of query being put to the system. In order to shorten the IR system's reaction time, we must focus on the kind and size of the corpus, the type of index, the type of query, as well as the searching methodology. Now let's look at data structure deployment in the context of information retrieval and how data structure selection impacts retrieval system performance.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, January 2022
Impact Factor: 6.252

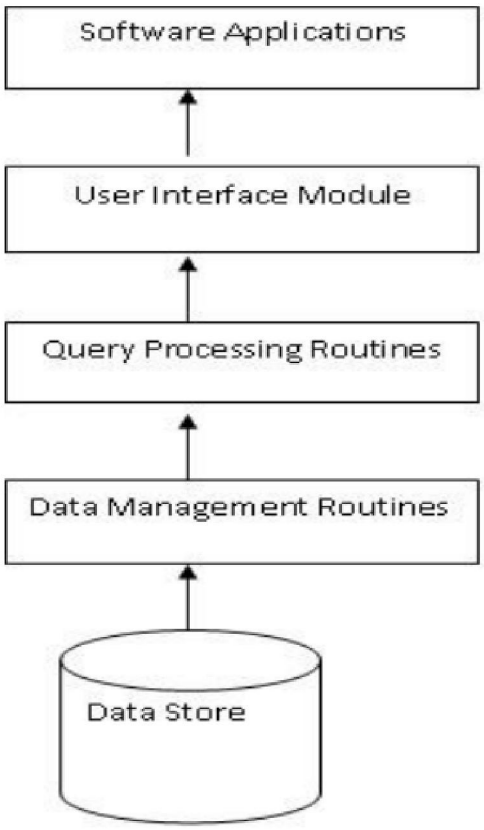

Figure 1. Architecture of a typical IR system

III. DATA STRUCTURES

The many methods used to store data in permanent memory are called data structures. Each application has a different data structure's function. The authors categorise the data in [3]. structures based on the function it serves. Storage structures are those that are primarily used for storing data, such as arrays, linked structures, and hash tables. Process-oriented data structures, which include stacks, queues, and priority queues, are another group of data structures that are used to process data. Some data structures are still unaccounted for because they do more than just store the data; they also allow the arrangement of the data to serve as a description of the data.

IV. RETRIEVAL OF INFORMATION TASKS

Answering a user's question is the main goal of information retrieval. In addition to finding answers to user inquiries, researchers are aiming to forecast the questions users may ask of a document corpus in the future. In [2], Fei Song and Bruce Croft calculated the odds of the user's query phrases being generated for each document in the corpus. Any information retrieval system's efficacy is determined by the user's feedback. In [4], the authors assess the effectiveness of information retrieval based on how well user preferences are satisfied.

V. STORAGE RETRIEVAL OF INFORMATION USING DATA STRUCTURES

For the purpose of obtaining the documents in response to a user query, information retrieval employs word focused indexing strategies. The hash function and hash tables are typically employed, however other indexing structures such as signature files [7], inverted files, etc. are also used. Key values and data items are connected via a hash data structure. The search key is mapped to a key value using a hash function. The bucket number to which the data item belongs is often indicated by the key value. A bucket is nothing more than a storage space. A hash table may be used as an in-memory data structure and is more efficient than the majority of array formats. Collisions are avoided by carefully choosing the hash functions. Comparing hashing to tree structures, which are better for range searches, it can be seen that hashing is more appropriate for equality searches. The hash functions produce an index that aids in locating the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, January 2022
Impact Factor: 6.252

pages that correspond to the user's query. For filtering, a hash file known as a signature file is employed. The filtering often pinpoints the pages that closely match the query. The hash function is used throughout the filtering process to generate a unique signature for each document. An inverted file has a hash file that has a list of sorted words, each of which is linked to its corresponding page by a series of pointers.

VI. DATA STRUCTURES THAT ARE PROCESS-ORIENTED FOR INFORMATION RETRIEVAL

A stack is a type of linear data structure that uses one end to store and retrieve data items. Information retrieval techniques match strings in suffix arrays using a stack. A graph is a type of data structure that has nodes and connected edges. It is one of the data structures with widespread use across several industries. It has been used to determine the connection between two computer network nodes or to determine the link between two data items or components. Graphs are employed in information retrieval to determine the connection between user queries and the documents in the corpus. The framework for idea networks utilised in fuzzy information retrieval is provided by graph structures. Every node symbolises a thought or a piece of writing. An edge is used in a concept network to link two distinct ideas, Ci, to a document, Di. The edge is labelled with a real number between zero and one, which denotes the fuzzy weighting assigned to the relationship. In web-based information retrieval, graphs are also utilised to provide relevance scores based on the relevance propagation in document graphs [9]. Collaborative filtering, categorization of the recovered documents, and unified link analysis are the additional applications of graphs in the field of information retrieval.

VII. DATA STRUCTURES FOR INFORMATION RETRIEVAL THAT ARE DESCRIPTIVE

A data structure known as a tree generates its subtrees with a node serving as the parent node and contains a data item as its root. Usually, the answer to a search tree is seen as a leaf. Depending on how they are set up and how they are traversed, trees come in many different varieties. A B tree is a binary search tree that furthermore possesses the ability to balance itself. The benefit of the B-tree is that searching takes just a logarithmic amount of time. A B+ tree is a self-balancing tree that has connected nodes that are used as pointers and can change its height. These pointers make it possible to effectively execute range searches in a B+ tree structure. In a digital tree, the right tree is traversed for a bit value of 1, while the left subtree is explored for a bit value of 0. Any search operation can often be thought of as the formation of a new node in a search tree. The index in the information retrieval process is implemented using binary tree structures like B trees and B+ trees. The inverted files are implemented using a B-tree. A Prefix B-tree reconstructs the tree each time it is searched [5]; it does not save the whole prefixes. The benefits of B-trees, digital search trees, and key compression approaches are all included in this system. Additionally, it lessens the processing burden that comes with compression methods. A string can be stored in a trie data structure starting at the root node and moving toward the leaf node. Figure 2 is a portion of [6], The authors illustrate how strings, an ape, an apple, an organ, and an organism are stored in a trie. A PAT tree is a binary tree structure used in the field of information retrieval. PAT is an acronym for PATRICIA, which stands for "Practical Algorithm to Retrieve Information Coded in Alphanumeric." Any route whose internal vertices all have exactly one child is compressed into a single edge by a PAT, a straightforward variation on a trie. a trie data structure with a radix of 2, which means that each node is a two-way (i.e., left versus right) branch and that each bit of the key is compared separately. Unlike the attempts, Patricia trees don't have any nodes with just one child. Every node has at least two offspring or is a leaf. This suggests right away that the ratio of internal (non-leaf) nodes to leaves is equal.\

VIII. CONCLUSION

Although we have developed an excellent basis to go ahead. Data Structures is a broad topic that includes more than simply stacks, queues, and linked lists. There are other more data structures, such as Maps, Hash Tables, Graphs, Trees, and so on. Each data format has benefits and disadvantages and should be adopted based on the demands of the application. A computer science student should be familiar with the fundamental data structures as well as the operations connected with them. Numerous of these data structures are incorporated into many high-level and object-oriented programming languages, such as C#, Java, and Python. As a result, understanding how things function behind the hood is critical. Dynamic storage allocation and reclamation are required for dynamic data structures. This can be

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, January 2022
Impact Factor: 6.252

done explicitly by the programmer or implicitly by a high-level language. It is critical to grasp the principles of storage management since these strategies have a substantial influence on programme behaviour. The basic concept is to preserve a pool of memory pieces that may be utilised to store dynamic data structure components as needed. When no longer required, allocated storage can be returned to the pool. It may be used and reused in this manner.

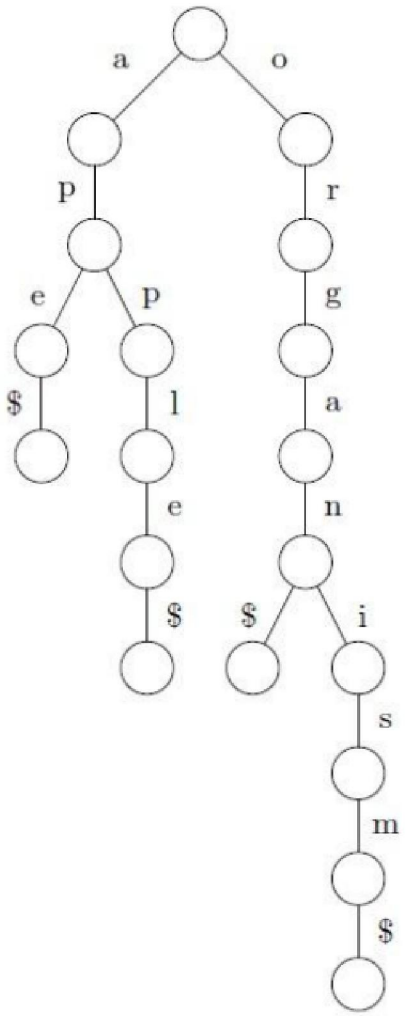

Figure 2: A Trie storing a string

Data structures called linear linked lists make it easier to traverse data elements quickly. It combines a number of data elements using pointers or references to the following data point. If it were a doubly linked list, we would have pointers to both the data item that comes before it and the data item that comes after it. Posting lists are implemented using linear linked lists. A posting list is a type of data structure used to keep track of documents that include a given phrase. Typically, a word dictionary is created, and for each phrase, a posting list is created with a list of documents that include that specific term.

REFERENCES

[1]. S. Ceri et al., Web Information Retrieval, Data-Centric Systems and Applications, DOI 10.1007/978-3-642-39314-3_2, © Springer Verlag Berlin Heidelberg 2013

[2]. Fei Song, W Bruce Croft. A general language model for information retrieval. Proceedings of the eighth international conference on Information and knowledge management (ACM) 1999/11/1. pp316-321

[3]. Falley. P "Categories of Data Structures", Journal of Computing Sciences in Colleges - Papers of the Fourteenth Annual CCSC Midwestern Conference and Papers of the Sixteenth Annual CCSC Rocky Mountain Conference. Volume 23 Issue 1, October 2007. PP. 147-153, 2007-10-01

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, January 2022
Impact Factor: 6.252

[4]. B. Zhou and Y. Yao Evaluating information retrieval system performance based on user preference JIIS, 34:227–248, 2010

[5]. Rudolf Bayer and Karl Undervaluer, "Prefix B Trees" ACM Transactions on Database Systems, Vol. 2, No. 1, March 1977, Pages 11-26.

[6]. Morin, Patrick. "Data Structures for Strings", chapter 7, March 2012.

[7]. Pogue, C. & Willett, P. (1987). Use of Text Signatures for Document Retrieval in a Highly Parallel Environment. Parallel Computing, 4, 259-268.

[8]. Nicholas. B Elkin, W.B Roces Raft, Retrieval Techniques, Annual Review of Information Science and Technology, Volume 22. 1987. Martha E. Williams, Editor Published for the American Society for Information Science (ASIS) by Elsevier Science Publishers