A Comparative Study of Desktop Search Algorithms

Harshita Shukla[1], Vinod Todwal[2]

[1]*Student, Department of Information Technology, Rajasthan College of Engineering for Women, Jaipur, India*
[2]*Assistant Professor, Dept. of Information Technology, Rajasthan College of Engg. for Women, Jaipur, India*

Abstract: **Desktop search algorithms are definitely becoming more and more important with rapid rise of memories and data size it is the necessity of any operating system to have a faster and more intelligent search algorithm. There have been many algorithms over the desktop searching few have been discussed in this paper.**

Keywords: **Desktop Search Algorithm, Data Packets, Big Hard Disks, Solid State Drives.**

1. Introduction

This paper demonstrates the detailed study of various papers over desktop search algorithms and represents what approach they made to identify and locate the files and the folders. There have been many search algorithms available with Operating Systems as an inbuilt feature but are usually slow and not as faster as they could be, also various companies like Google have launched many software with the inbuilt feature of data search but the they are usually of similar file extensions. So basically, there has been no generic search algorithm available as an inbuilt feature other than the Microsoft's Windows's inbuilt feature. However there have been many attempts made over the last few decades to reduce the search time for any particular file user has been looking for and the time has been definitely reduced as well, the search algorithms are definitely faster but they can be smart too. With the developments in the memory architectures there is definitely a possibility of searching data at more brisk rate. Earlier the systems were enabled with Parallel Advanced Technology Attachment Hard Drives (PATA) then they are replaced by Serial Advanced Technology Attachment Hard Drives (SATA) after that there are Solid State Hybrid Drives (SSHD) which are by default faster than the previous two, and above all those there are Solid State Drives which are superfast and they are based on transistor based memory architectures instead of using traditional magnetic disks architecture, however SSHDs uses both the concepts transistor based memories and magnetic disk based memories to ensure the large space with a very low cost.

2. Literature review

Chinta Someswara Rao et al [1] applied parallel string equivalent with JAVA Multithreading with multi core dispensation, and achieved a relative study on Knuth Morris Pratt, Boyer Moore and Brute force thread matching algorithms. For examining, gene sequence database are used which contains of absences of archives. From the test consequences it is exposed that the multi core dispensation is better associated to minor varieties. Lastly this planned parallel string matching with multi core dispensation is better associated to other sequential methods. Abdulwahab Ali Al-mazroi and Nur'Aini Abdul Rashid [2] planned a new hybrid algorithm called BRSS by uniting two algorithms, Berry- Ravindran and Skip Search. The hybrid algorithm determines improved character comparisons, number of efforts and examining time presentations in all the diverse data size and design lengths, consequently the proposed algorithm is useful for searching DNA, Protein and English text. This also showed that the request of the hybrid algorithm will lead to healthier searching and matching of the projects than the practice of one procedure as data is flattering more complex presently.

Saima Hasib, Mahak Motwani and Amit Saxena [3] deliberated the Aho-Corasick algorithm is best suitable for multiple patterns corresponding and it can be castoff in many application areas. The complexity of the procedure is linear in the measurement of the designs plus the time occupied of the examined text plus the quantity of output matches. It is originate to be good-looking in large statistics of keywords, meanwhile all keywords can be concurrently coordinated in one pass. Aho-Corasick delivers explanation to many actual world difficulties like Intrusion detection, Plagiarism detection, bioinformatics, digital forensic, text mining and numerous others. Aho-Corasick is one of the greatest creative algorithms in text mining.

Jorma Tarhio and Esko Ukkonen [4] planned a well-organized string-matching algorithm (named ACM) with compact memory as well as high worst-case presentation. By means of a magic numeral heuristic based on the Chinese Remainder Theorem, the planned ACM meaningfully reduces the memory necessity without passing complex procedures. Also, the dormancy of off-chip memory orientations is radically reduced. The planned ACM can be effortlessly applied in hardware and software. As an outcome, ACM permits cost-effective and effectual IDSs. Desktop search engines are castoff more and added in most contemporary operating systems. In overall, these search appliances extract metadata from

International Journal of Research in Engineering, Science and Management
Volume-2, Issue-1, January-2019
www.ijresm.com | ISSN (Online): 2581-5792

342

respectively file, and mass the metadata into an upturned index [5], a characteristic data construction used to provision keyword hunts. Data retrieval algorithms [6] are formerly used to regulate the best response to a enquiry.

Microsearch [7] is the primary that usages the upturned index to resolution the script retrieval in showy. It uses a buffer reserve to stock the metadata in reminiscence and flush the metadata to flash if the buffer reserve is occupied. In the upturned index building, it usages a per-page linked list and the whole head of the upturned index is immovable in the reminiscence, so the numeral of the slits in the inverted index is definite by the numeral of memory that can be cast-off in index creation and enquiry dispensation. A complement to Microsearch is PicoDBMS [8], a scaled depressed database for a smart card. Their project mechanisms well in a exact area and the operators are expected to be well qualified in the scheme. Microsearch on the other hand uses a search engine project which permits for formless inputs deprived of applying pre-specified qualities, and a natural language query border. The alteration amid the Microsearch and PicoDBMS can be summed up as the alteration amid a exploration engine and a database. Other embedded search schemes can be originate in sensor net literature like Macrohash[9]. Sensor scheme stocks a alike hardware construction, but they chiefly concerned with indexing and dispensation numeric data, and the enquiries are mostly variety query. In our project we must practice the text info.

3. Preview

An inverted index is commonly used in search engine systems to do full text search. A conventional inverted index named as record level inverted index has every slot on the upturned index resemble to a dissimilar term. Each slot supplies a pointer to a tilt of papers comprising that time, we call it text list. By matching a given query term with the upturned index, we can improve all the documents comprising that period. In this paper different approaches has been studied and indexing and segmentation based algorithms would be able to generate much better and faster results. Though from the above studied papers indexing can be straight algorithm with files getting indexed into the database, however segmenting them with sorting and then searching them will be a complex algorithm.

4. Conclusion

This paper presented a comparative study of desktop search algorithm.

References

[1] Ashish Prosad Gope, and Rabi Narayan Behera, "A Novel Pattern Matching Algorithm in Genome Sequence Analysis," (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4), 2014, 5450-5457.

[2] Abdulwahab Ali Al-Mazroi and Nur'aini Abdul Rashid, "A Fast Hybrid Algorithm for the Exact String Matching Problem," American Journal of Engineering and Applied Sciences 4 (1): 102-107, 2011.

[3] Ababneh Mohammad, OqeiliSaleh and Rawan A Abdeen, "Occurrences Algorithm for String Searching Based on Brute-Force Algorithm," Journal of Computer Science, 2(1): 82-85, 2006. K. Elissa, "Title of paper," unpublished.

[4] Pandiselvam. P, Marimuthu.T, and Lawrance. R, A Comparative Study on String Matching Algorithms of Biological Sequences," 2014.

[5] Justin Zobel and Alistair Moffat, Inverted Files for Text Search Engines. ACM Computing Surveys, Vol.38, No. 2, Article 6, 2006.

[6] M. Kobayashi and K. Takeda, "Information retrieval on the web," in ACM Computing Surveys, 2000.

[7] Chiu C. Tan, et.al., "Microsearch: When search engines meet small devices," Pervasive 2008, Sydney, Australia, May 19-22, 2008.

[8] P. Pucheral, et.al., "Pico DBMS: Scaling down database techniques for the smartcard," VLDB 2001.

[9] S. Lin, and D. Z. Yazti, et.al., "Efficient indexing data structures for flashbased sensor devices," in Transactions on Storage, Volume 2 Issue 4, pp. 468-450.