International Journal of Scientific Research and Engineering Development-– Volume 5 Issue 4, July- Aug 2022

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Deletion of duplicate Files & Images using Hashing Algorithm

Nagaveni B Nimbal
(Dept: CSE, College: K S School of Engineering and Management,Place : Bangalore
Email: nagaveni@kssem.edu.in)

----------**********----------

Abstract:

In computer hard drive is one of the core component. There is possibility to have same files on a same or different directory, searching for the same file in each directory is very difficult and take a long time. Duplicate File Searcher and Remover application is able to resolve. It will also be able to find out the same file that is located in a directory in hard drive.

These files are a mixture of important and unimportant files. Thus, it becomes infeasible for a user to organize and review these bulks of files and may lead to storage inefficiencies. So a solution is presented on this issue of inefficient storage. This issue is addressed by providing a user-friendly interface to users involving categorization of files into number of categories. The file identifies the redundant files in a folder, ,if the same files are located twice or more, then it directly deletes the redundant files from the system, without storing it in a recycle bin.

Keywords ***—Deletion ,Duplicate files, Redudancy, Directory, Files, Images, Audio, Video.***

----------**********----------

I. INTRODUCTION

While managing and performing file operations on computer or on other storage devices, many duplicate files with a considerable size may be gathered there. Accumulation of these digital junk levels can be a primary cause for shortage of storage space and decrease in computer performance.

Therefore, you need to search and erase duplicate files from computer hard drive. If talk especially about computer users, they should know how the access of duplicate files can affect their job. We all are aware of importance of RAM (Random Access Memory) in a computer. Duplicate file scans your hard drive for unnecessary duplicated files and help you remove them, freeing up space. Here are our picks for the best duplicate file finders, whether you're looking for something easy to use.

Data de-duplication is performed on file level and block level.The file level deduplication approach examines the operation of files on the basis of multiple aspects like index, name, time-stamp, etc.

These chunks are analyzed and compared to other chunks by using different hash algorithms. The unique hash value generated for each file and is compared with the different files. If a similar hash value is found, then it is considered as repeated data.

II. RELATED WORK

Deduplication has the ability to effectively manage storage allocation, significant cost savings as there is no need to buy extra storage space. Removal of duplicate data sustains network optimization. There is an enormous drop in both power and physical space requirements.

Finding and removing duplicate photos from PC is a more complicated task than finding a needle in a haystack. Also, these identical or similar-looking images tend to get piled up with time, clutter your photo library and consume up to GBs of disk space in your PC. That's why the fastest and safest

solution to find and delete duplicate images is to use the best duplicate photo finder and remover software.Since it is a really annoying and time-consuming task to manually scan and find duplicate photos from a huge collection of albums, therefore there is a definite need of getting a dedicated duplicate photo cleaner and remover tool that can automatically find and remove duplicate photos present in your system.These programs are renowned duplicate photos cleaners and can help you find and get rid of every kind of junk and duplicate photos that are causing your Windows PC to run slowly and adversely affecting its performance.To make your job much easier, we've handpicked some of the best duplicate photo finder and cleaner software available in the town to keep your system and photo gallery optimized. Our next segment focuses on the same.

III. DATA SET

Two datasets have been used for experimenting the currentmethod.Dataset1consists oflisting of all the files in a directory.

Fig. 1.Samples from Dataset 1 with listing files

Dataset2consistsoforginal files after removing all the redudancies in the directory.

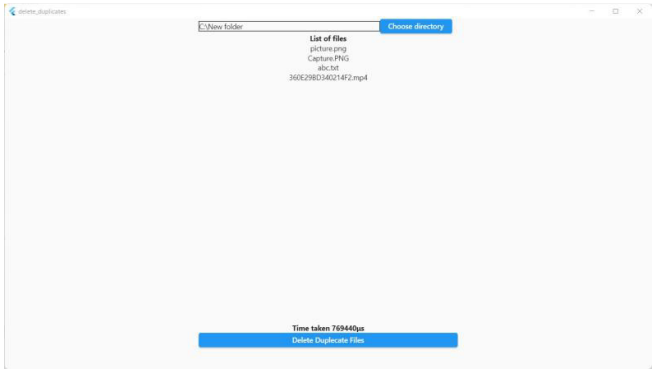

Fig. 2.Samples from Dataset 2 after deletion of files

IV. INCORPORATEDPACKAGES

A. Dart

Darts is open source and available here. You can install it in your favourite Python environment as follows: The basic data type in Darts is TimeSeries, which represents a multivariate (and possibly probabilistic) time series. It can be very easily built, for example from a Pandas DataFrame.

B. File Picker

The File Picker allows users to access various repositories. Repositories in CCLE enable users to upload files, access previously uploaded files and to easily bring content into CCLE from external repositories, such as Dropbox or Google Drive.A File Picker displays the information for orienting the users and to provide a consistent experience when users open or save files. That information includes - A tree of locations that the user can browse to.

C. Crypto

crypto provides a simple interface to symmetric Gnu Privacy Guard (gpg) encryption and decryption for one or more files on Unix and Linux platforms. It runs on top of gpg and requires a gpg install on your system. Encryption is performed with the AES256 cipher algorithm.

D. DATETIME

datetime module supplies classes for manipulating dates and times.While date and time arithmetic is supported, the focus of the implementation is on efficient attribute extraction for output formatting and manipulation.Date and time objects may be categorized as "aware" or "naive" depending on whether or not they include timezone information.

V. THE PROPOSED METHOD

Deduplication has the ability to effectively manage storage allocation, significant cost savings as there is no need to buy extra storage space. Removal of duplicate data sustains network optimization. There is an enormous drop in both power and physical space requirements. These helps to aid in removal of duplicate data by using hashing techniques and improve the efficiency. Hashing is so commonly used in computing that one might expect hash functions to be well understood, and that choosing a suitable function should not be difficult. The results of investigations into the performance of some widely used hashing algorithms are presented and it is shown that some of these algorithms are far from optimal.

Recommendations are made for choosing a hashing algorithm and measuring its performance.Hashing algorithm is a mathematical algorithm that converts an input data array of a certain type and arbitrary length to an output bit string of a fixed length. Hashing algorithms take any input and convert it to a uniform message by using a hashing table. Deletion of duplicate files & images using hashing algorithm. Duplicate files create exact copies of the original files with different names in windows. These duplicate files clutter up your drive with countless unwanted files. Duplicate files unnecessarily consume space in your computer and slow down the PC's performance. It makes sifting through files difficult and organizing data inconvenient, which might frustrate you.

Duplicate files enter through various ways into Windows and deteriorate its performance. If you regularly store a lot of data on your system, the chances of storing multiple copies of the file increase significantly. Another reason which can lead to duplicate files is multiple backups after merging drives and folders. It may happen again and again due to our ignorance.The multiple copies of duplicate files occupy a large section of storage that can make you run out of storage space. It slows down the system thus increasing the difficulty of navigating the original version of the files. Therefore, it is necessary to find duplicate files in Windows and get rid of them. There are various tools available in the market to help you find and delete duplicate files on your system. So in this article, we will show you the
three most effective methods to remove duplicate files.

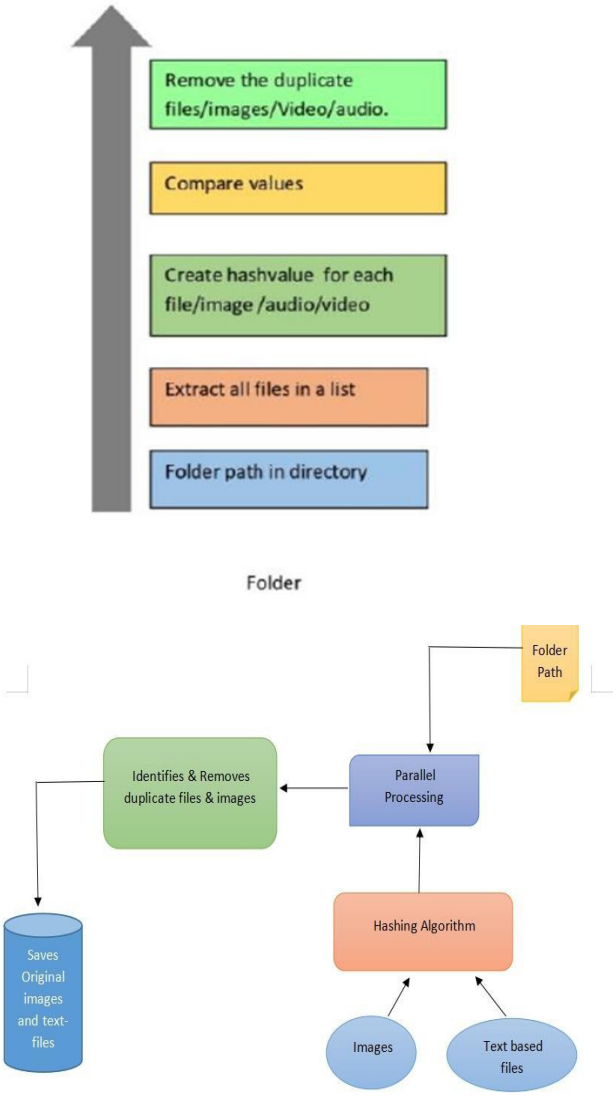

Fig. 1Proposed system

VI.RESULT AND ANALYSIS

Deletion of duplicate files,images,audio & video all original files will remain in the

directory,deleting all the duplicate files in the directory.We will get to know how much time it's taken to delete the files in the directory.System will free from the storage.

IV. CONCLUSIONS

In this paper, we briefly explained the deletion of duplicate files,images,audio,video We were able to deletete duplicate files and images using hashing algorithm. and also able to delete the duplicate files even if the file names are renamed. User interface design is implemented for removing of redundancy of files in a system.It has been reduces the storage areas of redundancy files in a system.

REFERENCES

[1] Ammar Asaad,Ali adil Yassin,"A New Scheme for Removing Duplicate Files from Smart Mobile Devices" August 2019 Cihan University - Erbil Scientific Journal.
[2] Ekta Thorat,Lekha Sonawane, Duplicate File Searcher and Remover, International Journal of Advance Engineering and Research Development Volume 4, Issue 4, April -2017.
[3] A Sachdeva, R Kapoor, A Sharma, A Mishra Categorical Classification and Deletion of Spam Images on Smartphones Using Image Processing and Machine Learning 2017 International Conference on Machine Learning and Data Science, p. 23 - 30 Posted: 2017.
[4] Dhanshree Wadile,Manisha Sonawane," Classification and Spam Image Detection in Smartphones,2nd International Conference on Advances in Science & Technology (ICAST) 2019 on 8th, 9th April 2019.
[5] O. A. FESTUS, "Data finding, sharing and duplication removal in the cloud using file checksum algorithm."
[6] E. Manogar and S. Abirami, "A study on data deduplication techniques for optimized storage," in 2014 Sixth International Conference on Advanced Computing (ICoAC). IEEE, 2014, pp. 161–166
[7] Abhishek Kadam, Bhagyshree Gawade, jidgnesh Sanke Duplicate: Redundant File Searcher and Remover by Sumita Chandak.

