TagFS - A Tag Based Filesystem

Catalina Macalet
Computer Science Department
Politehnica University of Bucharest
Email: catalina.macalet@cti.pub.ro

Eugen Hristev
Computer Science Department
Politehnica University of Bucharest
Email: eugen.hristev@cti.pub.ro

Mihai Dinu
Computer Science Department
Politehnica University of Bucharest
Email: mihai.dinu@cti.pub.ro

Sorin Dumitru
Computer Science Department
Politehnica University of Bucharest
Email: sorin.dumitru@cti.pub.ro

Abstract—**File systems are an integral part of every operating system. Because of the high capacity of modern hard drives file systems need a better way of organizing and accessing data.**

TagFS implements a tag-based filesystem in Linux which offers support for tagging files and browsing files by tags.

Index Terms—**Tags, file systems, VFS**

I. INTRODUCTION

In most operating systems the files are hierarchically organized. This means that there usually is a starting point, or parent directory. In Windows based systems there are multiple starting points based on the physical hard drive partitions. In Unix-like systems, there is a single root drive with different mount points available for users to add or remove subtrees from different drives, partitions, etc. In these filesystems a user organizes related data by storing it in the same folder but say that a user, Bob, has two separated folders one for storing photos taken in the mountains (Mountain-pics) and one for storing photos in which a certain person appears (Alice-pics). Two questions arise, one, where should Bob store a picture taken in the mountains in which Alice appears and two, how could Bob find the pictures taken in the mountains in which Alice appears. For current filesystems the answer to the first question might be storing the photo in either folder and in the second one creating a link to this photo or, store it in both folders. The answer to the second one could be naming the photo in such a way that retrieving them based on the previously stated criteria would work. TagFS file system aims to bring a different approach, based on tags rather than hierarchical system that is rooted for a long time in modern operating system. For the above example, for a TagFS filesystem the answer to both questions would be adding tags to photos (<mountains><Alice>) and then search for files that contain these tags. The question of where to store a specific photo would not be that important anymore. A pure tag file system is difficult to implement starting from zero, so we tried to adapt the current file system in Linux to support tags and see how the two systems can coexist on an end-user machine. A tag file system should be able to organize files, data on the disk regardless of hierarchical logical approach. The position of the files on the disk is irrellevant and completely transparent to the user. The file system should be able to put files on disks and simply recover them on demand based on tags requests. In our approach, logical directory based organization and file tags coexist, in order to see how the two systems can fit and how the user can use alternatives for searching and clustering the information it has. We implemented a tag layer in the Linux Virtual File System and tested how this impacts the regular user, we added posibilities for the user to play with tags: add, delete, search.

II. STATE OF THE ART

The idea of tagging files in order to access them in an easier fashion is not a new one and various attempts to implement solutions have been made.

A. Nepomuk-KDE

Nepomuk-KDE is an implementation of Nepomuk which has been integrated with KDE and that allows adding metadata to items stored on a computer and making queries based on that metadata. Dolphin KDE file manager allows adding and removing basic metadata to files, such as comments and tags.

B. Other tag-fs

existing kernel implementations for tag based filesystems.

III. TAGFS

A short description of how tagfs works.

A. Storage

B. VFS Hooks

IV. CONCLUSION

The conclusion goes here.

Possible future work: In a pure tag file system, the disk mechanism could be improved in the following way: We know that tags can be added to some files, we have no hierarchical structure of the files. This way we can find blocks of files based on tags which could reduce disk fragmentation. Clustering tag data can give insight on how much space there is required of a certain tag type files and how accessible this should be to the user. This could lower the external fragmentation of the disk if

properly used. However more tests should be done regarding this problem.
    -graphical if

REFERENCES

[1] Stephan Bloehdorn and Max Volkel,Tagfs-tag semantics for hierarchical file systems,
[2] Yuan-Liang Tai1, Shang-Rong Tsai, Guang-Hung Huang, Chia-Ming Lee, Lian-Jou Tsai, Kuo-Feng Ssu and Shou-Jen Wey, A Label-Based File System