

Faceted Metadata Based Searching and Browsing on Home Computer File Systems

Ishita Verma¹, Lorrie Faith Cranor², Kami Vaniea²

¹Indiana University Purdue University Indianapolis
Indianapolis, IN 46212 USA
iverma@iupui.edu

²Carnegie Mellon University
Pittsburgh, PA 15217 USA
lorrie, kami@cmu.edu

ABSTRACT

Traditional file system environments use a hierarchical directory system to organize and store files. However, such a model is not necessarily natural to end users, especially in the home environment where non-technical users do not think in terms of file hierarchies. Also in a scenario where home devices are connected to each other in an automatic file sharing network, the volume of files increase, making it difficult for users to set and recall file locations. Instead, we believe, end users are interested in locating files based on metadata information such as the type of file, who made it, when it was made, and what it was tagged as. In this research we developed an interface which allows users to browse files solely by their metadata. This paper presents the results of a user study on the effectiveness of using the faceted metadata based interface for file searching and browsing in a home environment. Our research shows that users are able to locate files, using our Faceted Metadata Browser, faster and with fewer mistakes than when using the Windows XP file explorer. Participants also expressed their liking for the new interface, largely due to the advantage of having greater number of ways in ways of reaching to a given file.

Categories and Subject Descriptors

H5.2. Information interfaces and presentation: User interfaces.

Keywords

Perspective, Faceted Metadata interface, data accessibility and user interface.

1. INTRODUCTION

Traditionally home computers users are accustomed to storing and accessing their files from file trees and searching for file names or textual content. This method is well practiced and effective, mainly due to the limited amount of files stored on personal computers and owing to the owner having some level of familiarity with her/his files.

Someday devices in a home will be linked allowing users to share files between computers seamlessly. Such a system would not only increase the volume of local files that can be accessed by one device, it will need to pass through hierarchical file locations for devices to automatically move files. Such a system would require a different approach than file hierarchy, and in this paper we proposed Faceted Metadata based file navigation.

Faceted Metadata based searching and browsing is rapidly becoming common on the web. Online web retailers, such as Ebay, Overstock and Amazon all allow their users to browse the store using meta data. This includes the multiple fields search forms of online library

or travel booking websites as well as cascading tag filters on online shopping websites.

Web search engines are another form of Metadata bases searching; they index the content of websites and use that data to assist users in their searches. Websites such as Google even use metadata about a page to effect the page's ranking or how it appears in the search results.

More recently, such faceted metadata based browsing has been added to online image collection, databases and even search engines such as Bing.

Despite increasing use of faceted metadata based browsing on the web, the idea has not been applied to devices for browsing files present on the system offline. More specifically, it is not yet used on personal home computers for accessing files in the system.

In this work we discuss a faceted meta data based user interface built on top of a distributed file system called Perspective[1]. The Perspective system was developed by The Parallel Data Lab at Carnegie Mellon for use in the home environment. In Perspective, the concept is that users can define and modify metadata on any given file, and locate files by filtering on files within all the metadata on the system, completely doing away with file hierarchies. This paper, discusses the design, usability and user reaction to an interface which used faceted metadata file browsing on a local file system.

We now describe the Perspective system which highlights the backend of a functional faceted metadata based file system.

2. OVERVIEW OF PERSPECTIVE

Perspective is a decentralized distributed storage file system designed for home users. It was developed at Carnegie Mellon Parallel Data Lab with the goal of making home storage more robust and accessible to users. In its present version, it enables all devices on a home network to store and access any data stored on any other device on the same network. It uses semantic rules to allow users to view and make decisions about what data is stored where.

The Perspective prototype is implemented in C++ and runs at user-level using FUSE [2] to connect with the system. It currently runs on both Linux and Macintosh OS X. Perspective stores file data in a repository on the machine's local file system and metadata in a SQLite database with an XML wrapper [1].

3. METADATA

Here we define and illustrate the notion of faceted metadata. "Data about data are referred to as metadata." [3]. Metadata on files is its attributes such as name, size, file type, length, ownership, and also user given descriptive information about the context, quality and condition, or characteristics of the data [4].

3.1 Faceted Navigation

Faceted navigation is to access a collection of files represented under multiple classification [5], here by their metadata. A facet comprises "clearly defined, mutually exclusive, and collectively exhaustive aspects, properties or characteristics of a class or specific subject" [6].

By assigning a faceted classification on a file using their metadata, there is no single

taxonomic order for reaching the file. Hence it can be browsed or searched using any of its metadata in any order. This allows user to browse or explore by filtering all the files by metadata keywords.

4. INTERFACE DESIGN

The design for the interface was inspired by the interface used by Amazon.com, Bing.com, Windows File Browser and the Flamenco Project by UC Berkeley [7]. This led us to keep separate features on the interface which would help in browsing, filtering and searching respectively (see figure 1.1 and 1.2).

For browsing, the facets were displayed as menu buttons which contained dropdown checkboxes containing all the facet values under it. These menus were located on the left end of the column and they were ordered alphabetically for the user to locate them easily. Most facets were the attributes that the system gives to every file such as date created and size. However there were also some user defined facets such as Author, Album, File Type, etc. The checkboxes or facet values that dropped down from the facet menus were static and non-changing.

To help navigate through the facets, there was a search bar on top of the facet menus which cleared out the facet menu area and displayed only those facets whose names contained the string of character being typed in the bar.

For filtering the search results, the bottom of the interface contained a list of related facets which dynamically updated with every facet value chosen to show what other facets can be used to filter the files. Clicking on any of the

facet names in the related facet area the facet menu area would clear out all menus and show only the facet menu whose name was clicked.

The search results area (in blue and white striped background) was given most of the interface space. This area showed a list of all the files that matched the currently selected facets. When no facets were selected no files were shown in this space.

For both browsing and filtering, there was a panel on top of the search result area. Every time a facet value was chosen, i.e. a query was passed a button with a small red "X" appeared on this panel with the facet value name. Hence this panel showed all present facet value queries in form of button. Clicking on the button would deselect the facet value and remove the query from filtering the result.

For searching by keyword, the top right hand corner contained a search bar which would sort and display files when searched for their content (in contrast to browsed by their metadata).

5. METHODOLOGY

The pilot study tested the Experimental Faceted File Browser interface against the Control Windows XP File Browser and Search Window interface. Windows XP was chosen as a control since it is the interface most often used in homes. To make this a fair comparison the Windows file browsing window was changed to "Detailed" view and extra columns were added so that relevant meta data was visible. The same meta data was displayed in Windows as was displayed in the Faceted Browsing interface.

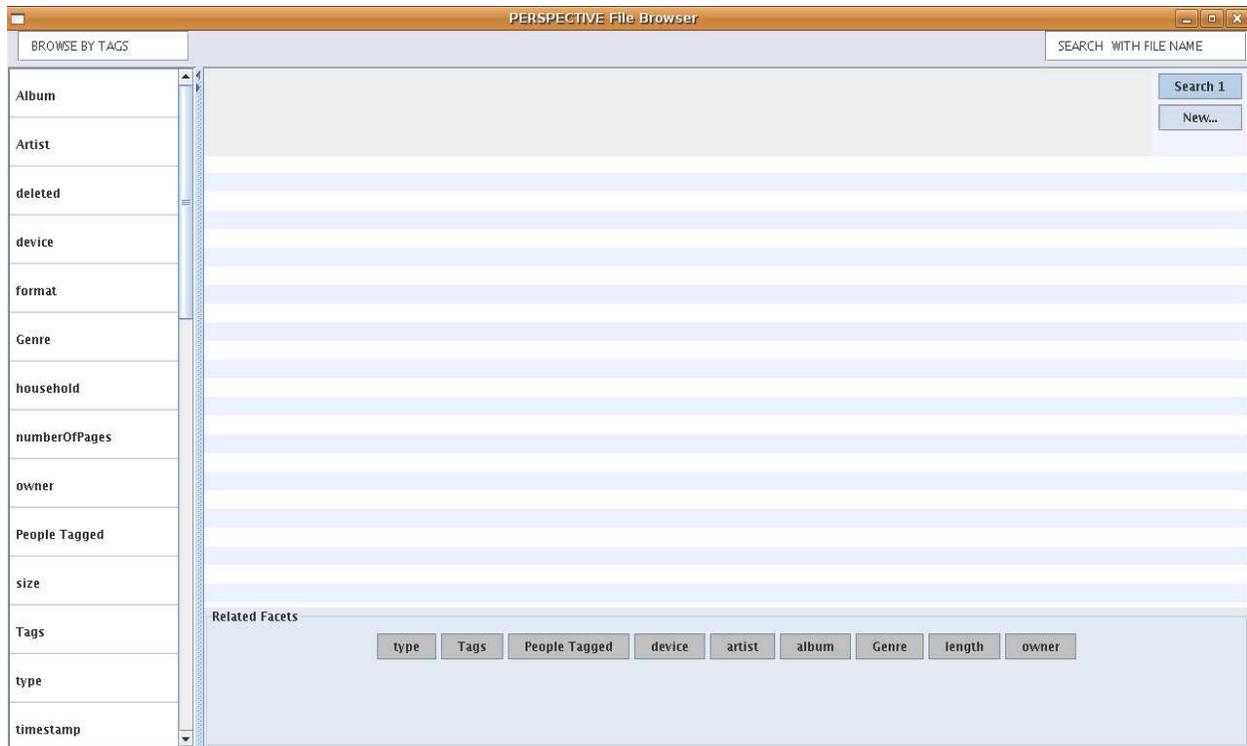


Fig 1.1 Default interface view

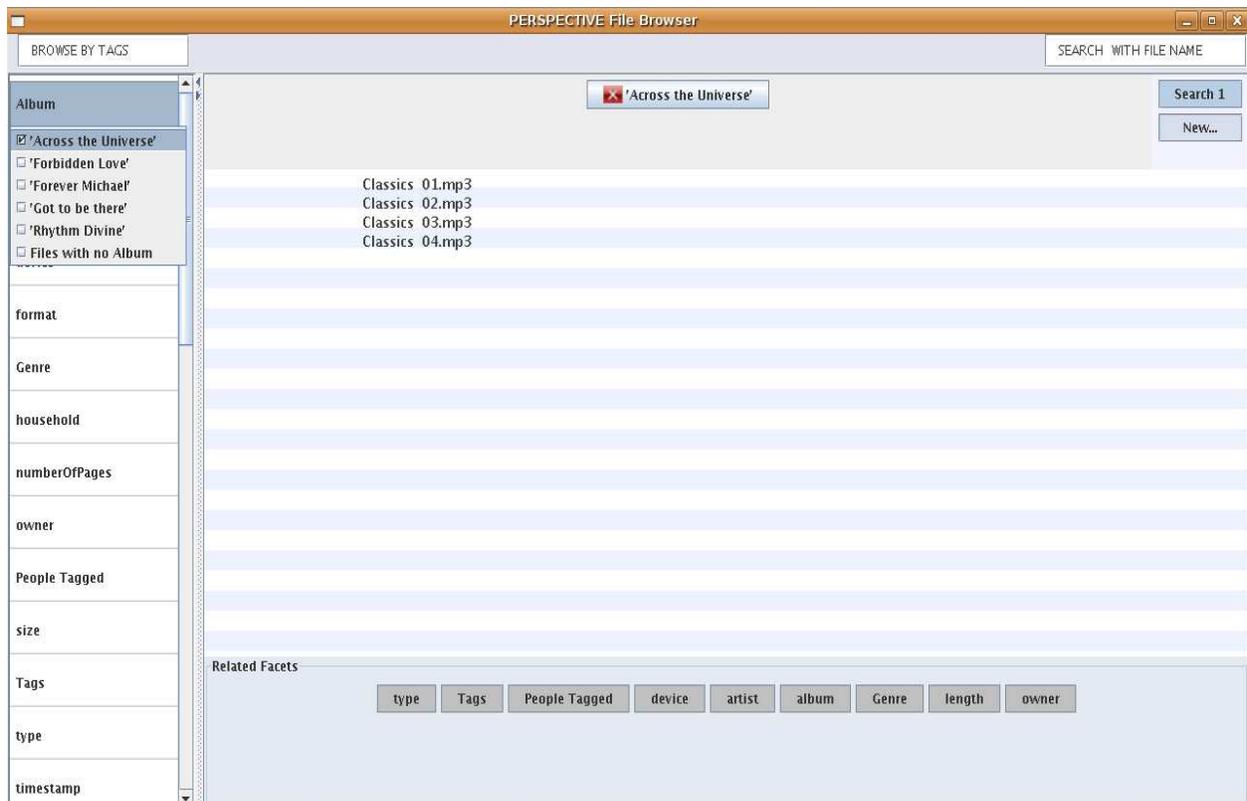


Fig 1.2 Interface showing one query selection.

The pilot study was conducted with six technical experts recruited from a computing laboratory. All six participants used both the Experimental section and the Control section.

The users in the study were in the 25 to 35 year old age range and worked on computers an average of 7 hours every day. The participants had previously worked on either Windows or Macintosh or Linux interface for a minimum of 10 years, but had never seen the Faceted Metadata interface before.

6. STUDY DESIGN

Trial Run

A trial study was conducted with two participants to validate the study setup and test the interface. After the trial we made some modifications to the task scenarios and the interfaces mainly so that the test users take a suitable minimum accurately recordable time to complete the tasks.

Tutorial

For both Windows and for the new interface the participant was walked through the various features of the interface by the experimenter so that the user was aware of all the functionality. For the experimental interface the user was made aware of the list of facets, the file list, the list of related facets, and the search bar. For the Windows interface users were shown how to get to the search interface and the extra columns that were added. While completing tasks users were given flexibility to choose which features they want to employ.

There were no training tasks.

Tasks

The goal of the research study was to measure the quickness and easy with which participants can browse for the desired files. Participants performed the exact same tasks using the Windows interface and the Faceted Browsing interface.

Participants were given three scenarios, which were presented in pre-determined order. The scenarios consisted to one to five related tasks. Each task asked the user to browse for specific audio, image and document files and filter or browse to only those files whose attributes match the given criteria.

The following is an excerpt of two tasks from one of the scenarios:

There were many people in your 2008 summer family vacation. But you are interested in pulling up the pictures of your Aunt Bonnie.

Find all the Image files which have the Tag - Summer 2008 and has Aunt Bonnie as People Tagged.

Your friends think your Aunt Bonnie looks quite young for her age, in fact they are more impressed with her hip hair color. You want to show them how Aunt Bonnie has changed her hair color every summer.

Find all the Image files which have Aunt Bonnie as People Tagged.

Data Collection

The users were asked to read the task aloud and then Think Aloud during the task. When they completed a task they were asked to say "I found it" when they were done with a task. Their voice was recorded and we used the audio to analyze the qualitative data.

The users were also timed on how long they took on each task. Timing was done between the time the user stopped reading the task and

when they said “I found it” at the end of each task. Their voice record of how long each user took on each task was used to cross check the time recorded during the session. This, along with the accuracy of their result gave us quantitative data.

7. PROCEDURE

The study session ran approximately 25 minutes with each participant. A study script was read out the participants first giving them a brief impression of the concept being studied. They were also given general instruction on how to ‘think aloud’ and why not to pace or plan their thoughts. They were also notified that they were being audio recorded and that excerpts will not be used in publication.

The participants were given a brief tutorial in the beginning for interfaces where they could ask questions. In this the researcher pointed out the general characteristics of both interfaces.

Users were alternatively chosen to test either the control interface or the experimental interface first. All users completed same tasks on both interface in the same order.

Each task was written on a slip of paper and handed to the user individually only after the previous task slips were handed back. This was done to ensure that the user was only focusing on a single task at a time.

The researcher sat behind the user in the back of the room and was unavailable to answer questions during the tasks. However the users were asked to ask questions anyways so that concerns with the interfaces could be noted.

At the end of the session the researcher joined the users for a debriefing session. Participants were asked to talk about their comparative comfort in using each interface. They were also asked if they felt certain advantages or drawbacks between the two interfaces. In the end the participants were thanked for their time and help.

8. RESULTS

8.1 Quantitative

Time Taken

The data collected on the average time taken by all users for every task on both interfaces is shown in Fig 2.1. Overall, there was an improvement in filtering tasks when faceted metadata was used. The graph is divided into five sections, depending on which interface performed better.

For the first Senario, perspective users spent a while exploring the facets in the list to find the facet that matched their goals. Windows, on the other hand, users were asked to perform a simple operation with which they were likely familiar, of opening the My Music folder.

However for tasks 1.2 and 1.3 which required filtering the search results to show only relevant files, faceted metadata performed better.

Tasks 2.1 and 2.2 both required Windows users to simply identify and open a folder. Consequently Windows performed better on these tasks. Faceted Browsing users had to locate the appropriate type of meta data to find the same information and consequently took longer.

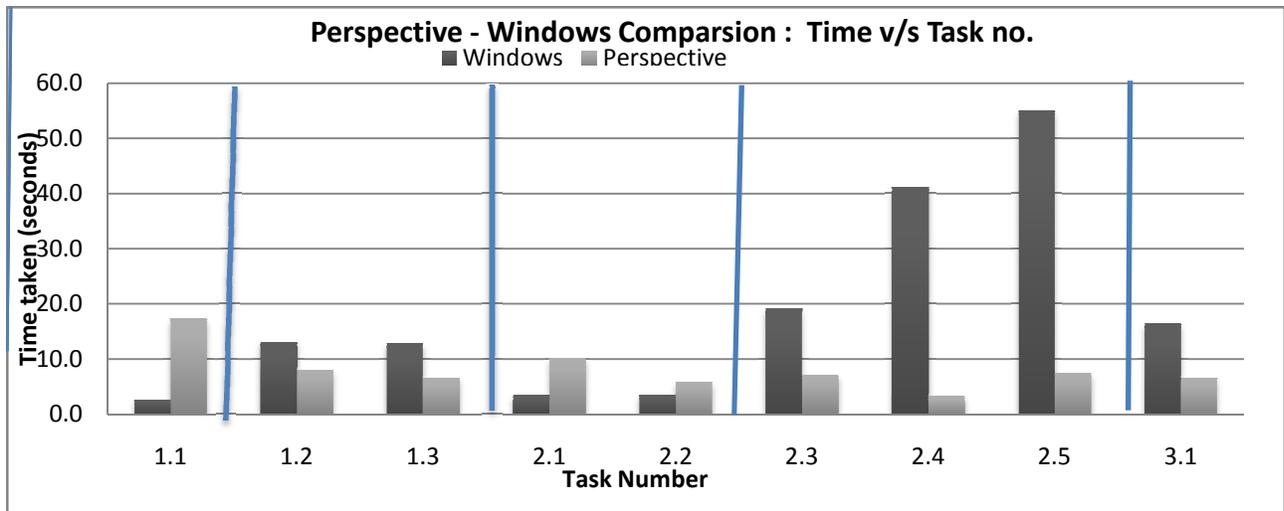


Fig 2.1 Time taken for performing each task on the Perspective and Windows File Browser interfaces

For subsequent filtering tasks 2.3, 2.4 and 2.5 Perspective users took only a fraction of the time that Windows user's required. Also, the time taken by Windows users increased as the complexity of filtering tasks increased.

The average time taken on the faceted metadata interfaced lessened over the tasks performed showing a learning effect.

Accuracy on Tasks

The faceted browsing interface performed better and with higher accuracy as it mainly avoided errors based on manually scanning and selecting files. The results are shown in Fig 2.2.

8.2 Qualitative

Information overload

For perspective, most users felt that there were too many tags showing on the tag panel. In contrast, there is no singular place in the windows file system which would show all the file system content and at any interface at any instance most of the information is hidden.

Cancelling previous queries

Most found it easier to filter results and cancel previous query in perspective file browser. In windows once within a directory path it was impossible to remove a query (i.e. folder name) in the middle of the path without losing everything below it.

	<i>Control: Windows</i>	<i>Experimental: Perspective</i>
<i>Average accuracy across all users (N=6)</i>	77.8%	88.9%
<i>Average accuracy across for users who made errors</i>	(N=3) 55.6%	(N=2) 66.7%

Fig 2.2 User accuracy in tasks performed

How much of the interface was utilized

In both perspective and Windows 4 out of 6 users felt comfortable with picking a single feature, such as search, and using it for the majority of the tasks. This led us to hypothesize that switching between searching and browsing functions is not standard to users.

9. DISCUSSION & SUMMARY

As the result for the user study it was found that Faceted Metadata navigation is a viable option for browsing files in local file systems.

Also faceted metadata approach improved user's ability at browsing and filtering, both by reducing time and by the ability to show all required files on a single interface view.

The interface design proved usable and effective within the context of the tasks.

In comparison with tradition file system, Faceted Metadata interface has some added functionality such as cancelling previous queries, giving and using user defined tags and combing searching and browsing together.

A more advanced interface and more rigorous user study may be able to identify any significant drawbacks in the design or concept.

10. RECOMMENDATIONS

There are many challenges that have to be addressed before faceted metadata can be proven to be effective in the home. Some of those challenges are:

- Allowing the user to add new metadata to files. Many files in an unstructured system will lack the file names which users currently make use of
- Automatic filtering facet names so that only useful facet names show up. What do users consider "useful?"

The interface also requires to dynamically update facets panel with every query selection such that:

- Facets containing no entry matching current query are removed.
- Show number of facets available which will also match the current query on top of the tab

The interface needs a way for user to customize tab panel and hiding some un-needed tags

References

- [1] Salmon, B. W. (2009). Putting Home Data Management into Perspective. Dept. of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh.
- [2] Filesystem in User Space Web Page. <http://fuse.sourceforge.net/>.
- [3] James Martin. (1987). Strategic Data Planning Methodologies, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
- [4] Metadata. In Wikipedia, the free encyclopedia. Retrieved August 18, 2009, from <http://en.wikipedia.org/wiki/Metadata>
- [5] Hearst, M. (2008). UIs for Faceted Navigation Recent Advances and Remaining Open Problems. In Proceedings of HCI08 Second Workshop on Human-Computer Interaction and Information Retrieval
- [6] Taylor, A. G. (1992). Introduction to Cataloging and Classification. 8th ed. Englewood, Colorado: Libraries Unlimited.
- [7] Hearst, M. A., Smalley, P., Chandler, C. (2006). Faceted Metadata for Information Architecture and Search. In Proceedings of CHI 2006 Montreal, Quebec, Canada.