How Do We Find Personal Files?: The Effect of OS, Presentation & Depth on File Navigation

Ofer Bergman
Bar-Ilan University
Ramat Gan, 52900 Israel
oferbergman@gmail.com

Steve Whittaker
University of California at Santa Cruz
1156 High Street, Santa Cruz, CA 95064, USA
stevejohnwhittaker@gmail

Mark Sanderson
RMIT University
GPO Box 2476, Melbourne 3001, Victoria, Australia
mark.sanderson@rmit.edu.au

Rafi Nachmias
Tel Aviv University
Ramat Aviv, Tel Aviv 69978, Israel
nachmias@post.tau.ac.il

Anand Ramamoorthy
Universiteit Ghent
9000, Ghent, Belgium
valvilraman@yahoo.co.in

ABSTRACT
Folder navigation is the main way that computer users retrieve their personal files. However we know surprisingly little about navigation, particularly about how it is affected by the operating system used, the interface presentation and the folder structure. To investigate this, we asked 289 participants to retrieve 1,109 of their own active files. We analyzed the 4,948 resulting retrieval steps, i.e. moves through the hierarchical folder tree. Results show: (a) significant differences in overall retrieval time between PC and Mac that arise from different organizational strategies rather than interface design; (b) the default Windows presentation is suboptimal – if changed, retrieval time could be reduced substantially and (c) contrary to our expectations, folder depth did not affect step duration. We discuss possible reasons for these results and suggest directions for future research.

Author Keywords
Personal information management, files, navigation

ACM Classification Keywords
H.5.2 User Interfaces: Evaluation/methodology, User-centered design.

INTRODUCTION
Most information retrieval research has focused on *public* data sources such as databases, libraries and the web, developing various theories and methods for retrieving such public information. Yet we all retrieve our *personal* files many times a day, predominantly using navigation. Personal file navigation (*navigation* for short) is a two-phase process [3]. First, users manually traverse their organizational hierarchy until they reach the folder in which the target file is stored. Second, they locate the file within that folder. Surprisingly little is known about this process; more specifically, how it is affected by the operating system used, the interface presentation and the folder depth. Our large scale quantitative study aims to deepen our understanding of navigation.

Industrial and research efforts have predominantly focused on search when addressing personal information retrieval. Much novel desktop search technology has been developed over the last few years; e.g., Google Desktop, Microsoft Windows Search, and Macintosh Spotlight. According to its advocates, desktop search promises to minimize users' organizational problems. Search reduces the need to manually organize personal information, which is automatically indexed by the search engine. Search has other potential advantages: it allows flexible and efficient ways to query one's personal information [6, 13]. Despite its promise, however, various studies still show a strong preference for navigation over search for accessing personal information when both are available [1, 5, 12, 15]. Moreover, the use of improved search engines has been shown to have little effect on this preference [3]. Bergman et al. [3] show that regardless of search engine quality, there is a strong preference for navigation.

It therefore seems that manual file organization and navigation remain critical Personal Information Management (PIM) behaviors that demand further study. While previous studies have looked at how we manually *organize* personal files [2, 7-12], less attention has been paid to retrieval; i.e., how people exploit these structures to *access* that information. To better understand navigation, we therefore conducted a large scale study testing for the effect of folder structure on file navigation [4]. In that study, we found folder structures to be shallow (files were retrieved from a mean depth of 2.86 folders), with small folders (a mean of 11.82 files per folder) containing many subfolders (M = 10.64). Navigation was largely successful and efficient with participants successfully accessing 94% of their files and taking, on average, 14.76 seconds to do so. Retrieval time and success depended on folder size and depth, indicating shallow but broad structures to be adaptive. Finally, a linear regression model identified an optimization point in this trade-off, leading us to recommend that users avoid storing more than 21 files per folder and create an additional level of subfolders instead.

CHI 2012, May 5-10, 2012, Austin, TX, USA.

However, there are still fundamental questions that can be asked about navigation, namely: does the *operating system* affect navigation – do Mac users navigate to files more quickly than PC users? There are also questions of *presentation*. There are multiple ways to present file information in a folder, e.g. details, icons, list, thumbnails, etc. Windows 7 currently presents the details view by default, but is this optimal for navigation in terms of retrieval time? Finally, in our earlier study, we did not test the effect of *folder depth* on step duration; i.e., the time it takes to navigate a single step down the folder tree. We assumed that folders that are higher in the file hierarchy would be retrieved more quickly because of greater familiarity. Higher levels should be more familiar because they are accessed more often. But is there empirical evidence for this assumption?

To the best of our knowledge, the effect of OS, presentation, and familiarity on file navigation have never been evaluated. In what follows, we report a large-scale study where our research questions were:

1. Does the operating system type affect retrieval success and retrieval time?
2. Does the folder's visual presentation affect retrieval time? If so, which of the views leads to shorter retrieval times?
3. Does folder depth affect step duration?

METHOD

To increase external validity, we collected data from large numbers of users who retrieved their own files on their computers. The requirement for lightweight, non-intrusive data collection led us to a procedure in which we recruited users and videotaped their screens as they accessed personal files from their own computers. We did not install software on people's machines to record organization and retrieval behaviors. Installation is error prone, and pilot interviews showed that users were concerned about its intrusiveness and potential implications for their privacy.

Research shows that users tend to access recent information items most frequently, regardless of whether these are files, web pages or emails [8-10]. We therefore videotaped participants navigating to files in their Recent Documents list; i.e., personal files that they had recently spontaneously retrieved and opened from their own computers, as part of their everyday computer use. There are a number of other important benefits to this approach. Focusing on recent files meant that users were trying to access files that we were confident were present on their disks and that were definitely retrievable by the user. It also allowed us to identify active files without having to manipulate or access participants' file systems, avoiding encroaching on their privacy.

Participants

Participants were 289 everyday computer users: 158 males, 131 females. The majority of participants were students and employees at a UK university. Participants' ages ranged from 16 to 64 years (M = 26.44, SD = 9.63). The majority of participants were Windows OS users (246: 181 XP, 62 Vista, 3 Windows 2000) and 43 used a Mac OS X operating system. We excluded 7 Linux users because their small number did not allow for reliable statistical comparisons.

Procedure

Participants used their own computers for the retrieval task. The tester printed out the participants' Recent Documents list, asking them to navigate to each file (the target) in that list in order. Participants were asked to click on the target file once but not open it. We did this to preserve users' privacy as target files might contain sensitive information. Participants were asked to close all open folders before each navigation task, so that all retrievals started from the desktop. Participants were asked to skip a file in the list if they had already navigated to that target folder during a previous access task. We did this to prevent access to these items being primed because that folder had already been accessed. We asked our participants to access only files saved on their computer and to avoid retrieving files on external drives and email attachments that hadn't been saved as files on their hard drive.

Retrievals and Time Measurements

Our study includes 1,109 valid retrievals. Of the initial set of 1,158 recorded retrievals, we excluded 4%, mostly because they were interrupted by external events such as phone calls or instant messenger alerts. Recordings of user interactions were made using a high definition digital video camera (with 1,080 horizontal scan lines) with a fixed rate of 25 frames per second, making each frame 40 milliseconds (0.04 seconds) long. We measured retrieval time by manually analyzing the videos frame-by-frame.

Retrieval Time was measured from the first mouse movement made by a participant in the navigation, until the moment when participants either clicked on the target file (in successful retrievals) or announced that they could not find it (indicating retrieval failure).

Step Duration – We use the term 'step' for each folder opened in the navigation process. Altogether, we measured 4,948 steps. Step duration was measured from the time a folder was opened until the time the user: (a) clicked on a subfolder to continue the navigation sequence, (b) reverted to a parent folder (if the relevant item was not found), (c) clicked on the target file, or (d) said, "I give up." Because step duration distribution was not normally distributed, we normalized it to Z scores for all statistical tests.

RESULTS

Does the operating system type affect retrieval success and retrieval time?

Table 1 shows the effect of operating system on retrieval success and time. A Mann-Whitney test showed no significant differences regarding retrieval success (row a); however, an independent sample t test indicates that Mac retrievals were significantly faster than PC retrievals (row b). As Table 1 shows, this difference cannot be explained by a difference in step duration (as no significant difference

was found in the step duration t test. See row c). Instead, this difference in retrieval time seems to be the result of different organizational strategies. Row d shows that Mac users stored their target files higher in the hierarchical tree than PC users. Mac users also had smaller folders (row e) containing more subfolders (row f). Our previous research [4] showed that folder depth and size are positively correlated with retrieval time. Therefore, it seems that Mac users use a slightly different storage strategy: keeping their folders shallower and smaller than PC users, and compensating for that with broader hierarchies.

	PC *N=246, 980 retrievals, 4,469 steps*	**Mac** *N= 43, 129 retrievals, 479 steps*	**Statistical Test**
a. Success rate	93%	96%	U=1.17 p>0.05
b. Retrieval time	17.27 (16.23)	12.56 (13.61)	t(1107)=3.15, p<0.001**
c. Step duration	3.3 (5.26)	3.13 (4)	t(4945)=.67, p=0.5
d. Folder depth	2.93 (1.87)	2.35 (1.63)	t(1030)=3.26, p<0.001**
e. Folder Size	25.05 (52.3)	19.04 (29.59)	t(3907)=2.24, p<0.001**
f. # of subfolders	8.85 15.67	10.73 40.42	t(2614)=1.97, p<0.05*

Table 1: The effect of OS on success rate, retrieval time, step duration, folder depth, folder size and number of subfolders in folder, showing Means and (Standard Deviations).

Does the folder's visual presentation affect retrieval time? If so, which view leads to shorter retrieval times?

There were 9 possible visual presentations of participant folders: 6 for Windows (Tiles, Icons, Details, Thumbnails, List and FilmStrip) and 3 for Mac (Icons, List and Columns). For a total of 4,209 steps, we were able to determine the visual presentation from the videos. Table 2 shows the effect of visual presentation on retrieval time.

Two one-way ANOVAs show a significant difference between the presentation groups for both PC $F(5,3792) = 19.08$, $p<0.001$ and Mac $F(2,408)=11.24$, $p<0.001$. Table 2 shows that *Icons* is the view that leads to the fastest step durations for both PC and Mac groups of presentation types. To test whether *Icons* led to significantly shorter step durations than the Windows default view (which was *Tiles* in Windows XP and *Details* in Windows 7), we conducted post hoc LSD tests. Both LSD comparisons were significant (and could not be explained by a difference in folder size or folder depth). On average, steps that used the *Icons* view instead of the *Details* view were 1.78 seconds faster. This means that if Microsoft were to change its default from *Details* to *Icons*, we might expect a reduction in average file retrieval time of 41%.

	No. of steps tested	**Step duration – *M* (*SD*)**
MS Icons	874	2.6 sec. (3.22 sec.)
MS Tiles	1,620	3.13 sec. (4.16 sec.)
MS List	313	3.21 sec. (4.73 sec.)
MS Details	609	4.38 sec. (8.88 sec.)
MS Thumbnails	345	5.45 sec. (7.91 sec.)
MS FilmStrip	37	5.5 sec. (5.38 sec.)
Mac Icons	229	2.22 sec. (2.01 sec.)
Mac List	123	3.57sec. (3.71 sec.)
Mac Columns	59	3.81 sec. (4.53 sec.)

Table 2: The effect of visual presentation on step duration.

Does folder depth affect step duration?

It is well known in cognitive psychology and HCI that familiarity reduces performance time [14], so we expected folder depth to increase step duration. Shallower folders should be more familiar as they are accessed more often. With an average of 10 subfolders in each folder, we can estimate that almost 100% of navigations pass through the main repository (e.g., My Documents), 10% pass through each of its subfolders, 1% of navigations involve its sub-subfolders, etc. To test this question, we conducted a regression analysis. Our data showed that step duration is positively correlated with folder size and folder size tends to decrease with depth [4]. We therefore partialled out the size effect using a stepwise regression. Results of the regression model ($R^2= 0.09$, $p<0.01$) show that folder depth did not significantly affect step duration when the effect of size was partialled out (see Table 3).

Factor	**Coefficient**	***SE***	***t***	***p***
Constant	2.05	0.07	27.61	<0.01
Size	0.03	0.002	16.78	<0.01
Depth	0.009	0.006	1.54	>0.05

Table 3: Regression model using size and depth to predict step duration.

DISCUSSION AND CONCLUSIONS

The results of our large-scale study are threefold:

(a) Mac users retrieve their files significantly faster than PC users. This difference seems to arise from the fact that Mac users deploy more sophisticated organizational strategies. Unlike PC users, they keep their files closer to the root directory by using small folders that branch sideways with a higher percentage of subfolders. The reasons for this behavioral

difference between PC and Mac users should be investigated in future research.

(b) Visual folder presentation affects the time it takes users to complete a step. The *Icons* view led to significantly faster retrieval than *Details*, which is currently the Windows default view. Our results indicate that changing the Windows default from *Details* to *Icons* would reduce retrieval time by 41%. However, there may be other usability considerations, e.g., the *Details* view adds information about the files and allows effortless transformation to chronological sorting.

(c) Folder depth had no effect on step duration, after partialling out the effect of folder size. This is a rather surprising result. A possible explanation is a ceiling effect (as these files were taken from the Recent Documents list making their retrieval paths highly primed). A more interesting possibility is that users see navigation paths as single conceptual entity, rather than consisting of a combination of elements broken down into each of the folders along the paths. This may be similar to physical navigation in which some (typically new) paths are remembered as a combination of turning points, while other (typically more familiar) paths may be remembered as a single entity (e.g. "the way home"). Future research could test whether file navigation involves a step-wise cognitive process ("OK I got this far, where to now?") or a holistic one, where the entire path is viewed as a single entity.

We also wanted to test for the effect of sorting on step durations; however, in only 2% of the steps was the sorting changed from the default alphabetical one, which was not enough for a valid statistical comparison.

Directions for future research include investigation of the effect of file presentation on step duration in a laboratory setting, tracking both eye and mouse movements. Controlled research involving eye tracking could investigate why some presentations lead to shorter duration time by manipulating file icon size, spatial order and the details of information added to them. These studies could in turn lead to informed design of new presentations with shorter retrieval time. Other research could investigate the effect of screen size on retrieval time. In our own work, we intend to further investigate the familiarity result by controlling folder familiarity, or using independent methods to measure it.

ACKNOWLEDGMENTS

This research was partially funded by the European Union Marie Curie Grant, PERG-GA-2009-248997.

REFERENCES

1. Barreau, D. K. and Nardi, B. A. Finding and reminding: File organization from the desktop. *SIGCHI Bulletin 27*, 3 (1995), 39-43.
2. Bergman, O., Beyth-Marom, R. and Nachmias, R. The project fragmentation problem in personal information management. In *Proc. CHI 2006*, ACM Press (2006), 271-274.
3. Bergman, O., Beyth-Marom, R., Nachmias, R., Gradovitch, N. and Whittaker, S. Improved search engines and navigation preference in personal information management. *ACM Transactions on Information Systems 26*, 4 (2008), 1-24.
4. Bergman, O., Whittaker, S., Sanderson, M., Nachmias, R. and Ramamoorthy, A. The effect of folder structure on personal file navigation. *Journal of the American Society for Information Science and Technology 61*, 12 (2010), 2426–2441.
5. Boardman, R. and Sasse, M. A. "Stuff goes into the computer and doesn't come out": A cross-tool study of personal information management In *Proc. CHI 2004*, ACM Press (2004), 583-590.
6. Cutrell, E., Dumais, S. T. and Teevan, J. Searching to eliminate personal information management. *Communications of the ACM, 49*, 1 (2006), 58-64.
7. Gonçalves, D. and Jorge, J. A. An empirical study of personal document spaces. In *Proc. DSV-IS'03* (2003), 46-60.
8. Hardof-Jaffe, S., Hershkovitz, A., Abu-Kishk, H., Bergman, O. and Nachmias, R. How do students organize personal information spaces? In *Proc. Educational Data Mining 2009* (2009), 250-258.
9. Hardof-Jaffe, S., Hershkovitz, A., Abu-Kishk, H., Bergman, O. and Nachmias, R. Personal information space organization strategies in learning context. *Special Issue of Journal of Digital Information on the Social and Psychological Aspects of Personal Information Management 10*, 5 (2009), 1-16.
10. Henderson, S. and Srinivasan, A. An empirical analysis of personal digital document structures. In *Proc. HCI International 2009* (2009), 394-403.
11. Jones, W., Phuwanartnurak, A. J., Gill, R. and Bruce, H. Don't take my folders away! Organizing personal information to get things done. In *Ext. abstracts CHI 2005*. ACM Press (2005), 1505-1508.
12. Kirk, D., Sellen, A., Rother, C. and Wood, K. Understanding photowork. In *Proc. CHI 2006*, ACM Press (2006), 761-770.
13. Russell, D. and Lawrence, S. Search everything. In W. Jones and J. Teevan (Eds.), *Personal information management*. University of Washington Press (2007), 153-166.
14. Shneiderman, B. and Plaisant, C. *Designing the user interface.* Addison-Wesley Publ. Co., Reading, 2010.
15. Teevan, J., Alvarado, C., Ackerman, M. S. and Karger, D. R. The perfect search engine is not enough: A study of orienteering behavior in directed search. In *Proc. CHI 2004*, ACM Press (2004), 415-422.