https://doi.org/10.48009/3_iis_2020_236-244

Issues in Information Systems
Volume 21, Issue 3 pp. 236-244, 2020

FILE MANAGEMENT: STUDENT KNOWLEDGE OF FILE TYPE EXTENSIONS

Paul Stephens, Bradley University, prs@fsmail.bradley.edu
Matthew McGowan, Bradley University, mmcgowan@fsmail.bradley.edu

ABSTRACT

This study looks at one small facet of file management, files extensions, in order to understand what current students know about file types and their associated applications. We examine the file management literature and identify themes and actions from previous research that relate to the use of file extensions. We develop and conduct a survey of undergraduate college students about file extensions. We find some differences based on which type of OS a student's personal computer has, but few differences based on their year in college.

Keywords: File Management (FM), file extensions, file types, personal information management (PIM)

INTRODUCTION

File management is a ubiquitous practice which may be taken for granted. The fact that it is universally practiced by students and working professionals on a regular basis should signify its importance to educators. Therefore, understanding what students know or need to learn should be the focus of pedagogical research. As academic professionals, we have noticed that the file management skills of our students seem to have degraded over the years. We have observed many instances of students not knowing where files are located or subsequently, how to find those files. Students routinely post incorrect versions of files for homework and sometimes exams. Students have submitted files such as HW1.xlsx.xlsx. On occasions when we get the chance to view their personal computers, we often find a disorganized mess. The purpose of this research is to examine how well today's students understand the fundamentals of file management. It is an important topic for students in their academic careers, and will likely be important in their professional careers. It is important to computer educators because they need to be able to help students improve their ability to manage personal knowledge. Do they understand that they can manage their personal files? Do they know how? Or, are most students simply "pilers": people who don't file their documents, they just let them pile up on their desktop or in the downloads folder (Henderson and Srinivasan, 2011)?

BACKGROUND AND LITERATURE REVIEW

Barreau (1995) defined a personal information system (PIM) as "an information system developed by or created for an individual for personal use in a work environment. It includes a person's methods and rules for acquiring the information which becomes part of the system, the mechanisms for organizing and storing the information, the rules and procedures for maintaining the system, the mechanisms for retrieval, and the procedures for producing the various outputs required." One of the personal information management tools available for use on a computer is the file manager. Prior to the time personal computers had graphical user interfaces (GUIs), users had to understand how to interact with the directory structure of the computer to know how to retrieve and store files. With the advent of GUI and the widespread adoption of Windows-based operating systems, most users still needed to get familiar with the storage and file retrieval structure of personal computers. Two or three decades ago, universities were providing this fundamental knowledge to students as it was necessary to their academic productivity. As the technology progressed, features like recent files, searching, and tagging became available. Students also have more options for where to store files: personal computer, personal flashdrive, mapping drives to network storage and cloud services are available (Hardof-Jaffee, et al., 2009). Some of these features were attempts to simplify file management. For example, there was a default download folder, there was a default folder for storing user work, the recent feature does not show where files are stored, only that they were recently accessed. The student need for specific training in operating systems and file management seemed unnecessary. Yet the number, variety and complexity of choices in how to acquire, store, manage, share and synchronize files has increased as well.

FM literature which focuses on user behavior describes how and why users practice file management. Underlying this research is the assumption that users understand operating systems and file management well enough to make

informed choices. This is not likely true for all users and more so for university students. Students are likely both ignorant and unorganized in varying degrees. Ignorance can be addressed much more easily than modifying student behavior. But, efforts in the classroom can attempt to address both. To address ignorance, we need to know where it exists in terms of the four file management themes described by Dinneen & Julien (2020); storing, organizing, retrieving and sharing. For this paper, we empirically examine a small part of file management; file extensions.

File Management Themes

To put this research in perspective, a brief overview of file management can provide the context for our exploration into file extensions. In their comprehensive literature review, Dinneen & Julien (2020) have laid out four file management themes but have also categorized the user actions within each theme. See table 1. Additionally, they address the tools used to direct these actions; hardware, operating system, and file management applications.

Table 1. FM Themes and Associated Actions (Dinneen & Julien, 2020)

FM Theme	Actions
Storing	creating, downloading, filing, naming, backing up files
Organizing	creating subfolders, moving and deleting files and folders
Retrieving	navigating, searching, tagging files and folders
Sharing	sending files, negotiating storage, collaborative organization, retrieval in shared space

Each of these actions and the tools that enable each action are ripe for investigation in attempting to capture the educational state of our current students. Students can practice file management and can choose to do so in an effective and efficient manner only when they are able to harness the knowledge and skills needed for such practice. Before we can identify what knowledge and skills we need to address as educators, we need to understand what students know when they come to us.

File Extensions

FM research has looked at file types or extensions, in terms of their numbers, for the most common types of files stored (Agrawal, Bolosky, Douceur & Lorch, 2007; Jensen, et al., 2010; Henderson, 2011; Dinneen & Julien, 2019). But, in terms of their usefulness in the practice of file management, little can be found. Blanc-Brude & Scapin (2007) found that of all the different attributes concerning a file that a user can recall, the file type (e.g. MS Word document or Adobe PDF) had, by far, the highest percentage of correct recollection at 93.3%. The correct recollection of the file name itself was only 25% and keywords were 32%. Clearly, a general knowledge of file extensions can help users with actions (Dinneen & Julien, 2020) related to: Storing, Organizing, Retrieving and Sharing (see Table 2).

Table 2. Example actions potentially impacted by file extensions

FM Theme	Actions
Storing	creating, naming
Organizing	moving
Retrieving	navigating, searching, tagging
Sharing	sending files

For example, if a person is looking for a file related to a project on franchising, they might have similarly named files for reports, presentations and cash flow analyses. Knowing both a keyword and file extension will narrow down the search and reduce the time to find a file. Therefore, if the person is looking for a file related to the franchising project and knows he or she is looking for a MS Excel file, then they can start searching based on both franchising and the .xlsx file extension. Additionally, file extensions have an impact on naming and renaming files. A user might use the word "report" as part of a file name when saving a file in a word processor. Thus, the

combination of the keyword report and the extension .docx helps the user find the file. Or perhaps the user has done some analysis in SPSS but this isn't software she uses very often and she can't remember the name of the file. Starting with a broad search for .sav files might be a quick place to start.

The relationship between a file extension and moving or sending a file isn't necessarily apparent to students. One issue we observe is when students attempt to move or upload files while the files are still open in an application. For example, when working with an MS Access database, a record locking file opens up when the database is opened. If a student looks at the folder in the file manager, they will see both the database, HW9.accdb and HW9.laccdb. The student will sometimes move or upload the locking file and not the database. If they are working in a temporary workspace, or a shared lab computer, they risk the loss of their work.

RESEARCH METHODOLOGY

File management research assumes a certain level of knowledge, skills and abilities of users. It examines file management practices in terms of their behavior or the usefulness of the FM tools themselves. We posit that a lack of knowledge in understanding file management exists amongst users. And, this is likely truer for less experienced users of technology, college students. This lack of knowledge alters user behavior and affects their use of file management tools.

This research attempts to assess what business students know about file extensions in a MS Windows OS environment. We developed a survey and gave it to students in a variety of management information system courses. The students represented various stages of academic careers from freshman to seniors. Additionally, the personal information management environment of each student, represented by their own personal computers (Mac OS or Windows) was captured as well. From this we can test a series of null hypotheses.

A quiz was used to test the knowledge of the student. Either the student answered a question correctly or they didn't. The outcome of interest is dichotomous or binary (e.g. correct/incorrect). From that we can calculate the proportion of students who answered the question correctly. Then we tested the proportional inequality between two independent comparison groups using the z-test for independent proportions (Holmes, Illowsky & Dean, 2017) where the null hypothesis is that the proportions are equal: Windows versus Apple users and upperclassmen (Juniors and Seniors) versus lowerclassmen (Freshmen and Sophomores). In two cases we also test a specific proportional claim (H3 and H6).

We wanted to see if students understood how to enable or disable file extensions using file manager in the Windows OS. We might expect students who owned Windows computers to be statistically more knowledgeable about this than students whose personal device was an Apple. But, this expectation is confounded by the fact that students are required to work in a Windows OS in computer labs and in the virtual desktop environment for their courses.

- H1: Students whose personal computer is based on the Windows OS will <u>not</u> know more about how to manage file extensions in a Windows OS than students whose personal computer is based on the Mac OS.
    - H0: $p_A - p_W = 0$
    - Ha: $p_A - p_W <$ or > 0

Regardless of their personal computer, students who have taken more classes and have worked in the Windows OS environment provided by the university are more likely to have developed file management skills.

- H2: Upperclassmen (Juniors and Seniors) will know more about how to manage file extensions in a Windows OS than lowerclassmen (Freshman and Sophomores).
    - H0: $p_L - p_U = 0$
    - Ha: $p_L - p_U <$ or > 0

And, we desired to see how well students understood file extensions by matching the file extension to the correct application. For business students, we expected that a significant majority would be able to recognize certain

common extensions associated with software they use in several courses: Adobe Acrobat, MS Office, Pictures, Text files and Web pages.

- H3: A majority of students will recognize common extensions associated with software they use in their courses.
  - H0: p=.50
  - Ha: p>.50

- H4: Students whose personal computer is based on the Windows OS will not correctly identify more common file extensions than students whose personal computer is based on the Mac OS.
  - H0: $p_A - p_W = 0$
  - Ha: $p_A - p_W <$ or > 0

- H5: Upperclassmen will correctly identify more common file extensions than lowerclassmen.
  - H0: $p_L - p_U = 0$
  - Ha: $p_L - p_U <$ or > 0

There are a variety of applications that students may encounter during their time at university or in other environments such as an internship. Yet, students will have much less experience using many of these applications and are much less likely to recognize their file extensions.

- H6: A minority of students will recognize uncommon extensions associated with applications they are unlikely to use on a regular basis.
  - H0: p=.50
  - Ha: p<.50

RESULTS

The survey was given to (138) students in six different MIS classes in February and early March of 2020. Of these students, 36%(50) use an Apple computer as their own personal machine and 64%(88) use a Windows PC. There were fifty-two upperclassmen(38%) and eighty-six lowerclassmen(62%).

With the default installation of either Windows or Mac operating systems, when using file explorer or finder, file extensions are turned off. In both instances, file extensions can also be turned on. Since our college of business has created a Windows environment for our computer labs and virtual desktop, we asked students about the default condition and specifically how they would go about turning files extensions on or off in a Windows environment.

A minority (43.5%) of all students knew that, by default, file extensions are turned off in the Windows operating system. There were no statistically significant differences between Apple (42.0%) and Windows (44.3%) users giving the correct answer (p=.76). The null hypothesis for H1 cannot be rejected. The data doesn't support any claim that differences in knowledge exist between the two groups.

A similar analysis was performed to test who knew how to manage (turn on or off) file extensions in the Windows OS. Only 26.8% could pick the correct method from four choices. The proportion of Apple students who picked the correct answer was (20.0%) and Windows students (30.7%). The null hypothesis for H1 cannot be rejected (p=.17). The data doesn't support any claim that differences in knowledge exist between Windows and Apple users.

Both groups were equally lacking in their knowledge of managing file extensions in a Windows OS environment. So, do students actually learn how to manage file extensions as they proceed through their academic courses?

There were no statistically significant differences between Upper (46.2%) and Lower (41.9%) users giving the correct answer. The null hypothesis for H2 cannot be rejected (p=.61) The data doesn't support that upperclassmen are any more likely to have the correct knowledge about the default setting for file extensions than lowerclassmen.

A similar analysis was performed to test who knew how to manage (turn on or off) file extensions in the Windows OS. The proportion of Lowerclassmen who picked the correct answer was (24.4%) and Upperclassmen (30.8%). The null hypothesis for H2 cannot be rejected (p=.42). The analysis does not support that upperclassmen have superior knowledge about how to turn on or off file extensions using Windows file explorer.

Fortunately, most students know that file extensions exist. Therefore, they can learn how to manage file extensions. What we show in the following hypotheses is that a majority of students know about and can recognize common file extensions. Yet, a noteworthy number of students are unable to demonstrate their understanding of common file extensions.

We gave students examples of files with extensions and asked them to identify the application with which they were associated. See Table 3.

When we examined the results from the common extensions, the majority of students were able to correctly match the extension with the application. The null hypothesis for H3 can be rejected (p<.05). A majority of students will recognize common extensions associated with software they use in their courses. See table 4. The exceptions were MS Access databases and Web Pages (p>.05).

Table 3. Applications and associated file extensions that were tested

Application	**File Extension(s) Tested**
Adobe Acrobat*	.pdf
Adobe Photoshop	.psd
Bitmap Picture	.bmp
Comma Separated Value	.csv
Compressed Archive*	.zip
Extensible Markup Language	.xml
Graphics Interchange Format	.gif
Hypertext Preprocessing Code	.php
Java Script	.js
Java Script Object Notification	.json
Joint Photographic Experts Group*	.jpg
Microsoft Access Database*	.accdb, .mdb
Microsoft Excel*	.xlsx, .xlsm
Microsoft PowerPoint*	.pptx, .ppam
Microsoft Word*	.docx, .dotm
Portable Network Graphic	.png
SPSS (Statiscal Package for the Social Sciences)	.sav
Structured Query Language	.sql
Tableau	.twb
Tableau Prep Builder Flow	.tfl
Text*	.txt
Web Page*	.htm
*Common applications for business students	

We did find some statistically significant differences (P<.05) between Apple and Windows users in three cases: Adobe Acrobat, Compressed Archive (.zip) and MS Access. In each case, the Apple user was proportionally less likely to choose the correct application. This may indicate that a student whose personal computer is an Apple is

less likely to create these types of files or lacks the software to create these types of files. Therefore, he or she is less likely to recognize the associated file extension. This rejects the null hypothesis for H4. In all other cases, no significant differences were found, thus the null hypothesis for H4 cannot be rejected.

Surprising to us, was that only 36.5% of the upperclassmen were able to recognize the MS Access database extension. Almost every one of those students had multiple classes in which that software has been used. For each common extension, there were no significant differences in results between upper and lower classmen. The null hypothesis for H5 is not rejected. This may indicate that students aren't appreciating the value of file extension knowledge and thus are not attempting to acquire such knowledge as they proceed through their academic program. And, any knowledge they do acquire early in their classes, is sometimes lost by the time they become upperclassmen.

Table 4. Percentage of students who correctly identified common application used in the college of business

		% Correct				
Application	Extension	All	Apple	Windows	Lower	Upper
Adobe Acrobat	.pdf	53.6%	**40.0%**	**61.4%**	53.5%	53.9%
Compressed Archive	.zip	56.5%	**44.0%**	**63.6%**	53.5%	61.5%
Joint Photographic Experts Group	.jpg	60.9%	62.0%	60.2%	59.3%	63.5%
MS Access Database	.accdb	42.0%	**32.0%**	**48.0%**	45.4%	36.5%
MS Excel	.xlsx	92.0%	96.0%	89.8%	91.9%	92.3%
MS PowerPoint	.pptx	88.4%	84.0%	90.9%	87.2%	90.4%
MS Word	.docx	84.0%	90.0%	80.7%	82.6%	86.5%
Text	.txt	96.4%	96.0%	96.6%	96.5%	96.2%
Web Page	.htm	40.6%	44.0%	38.6%	38.4%	44.2%
In Bold - statistically significant proportional differences (p<.05)						

Note that sometimes the extension can be intuited from the name of the application. For example, it isn't difficult to make the leap from .sql to structured query language, even if you don't know what SQL is and you have never used it. Additionally, some extensions related to common applications have fewer common extensions, like macro enabled files in MS Office. Another example is a discontinued file extension, .mdb, an MS Access file in an older format. The results from less common applications and extensions is shown in table 5. The null hypothesis for H6 can be rejected (p<.05). Thus, statistically supporting our assertion that a minority of students will recognize uncommon extensions associated with applications they are unlikely to use on a regular basis.

Table 5. Less Common Applications and Extensions

		% Correct				
Application	**Extension**	**All**	**Apple**	**Windows**	**Lower**	**Upper**
Adobe Photoshop	.psd	15.2%	10.0%	18.2%	15.1%	15.4%
Hypertext Preprocessing Code	.php	15.2%	20.0%	12.5%	14.0%	17.3%
MS Access Database	.mdb	61.6%	**72.0%**	**55.7%**	59.6%	61.6%
MS Excel	.xlsm	47.8%	40.0%	52.3%	51.1%	42.3%
MS PowerPoint	.ppam	5.8%	4.0%	6.8%	3.5%	9.6%
MS Word	.dotm	2.9%	4.0%	2.3%	3.5%	1.9%
SPSS	.sav	26.1%	26.0%	26.1%	24.4%	28.9%
In Bold - statistically significant proportional difference (p<.05)						

Interestingly, students were able to discern .mdb as Microsoft database and therefore assign it to the correct application. As opposed to the newer version of MS Access, .accdb, where only 42% could assign it correctly.

When file extensions change over time, it creates another level of understanding and knowledge that only comes with experience. And, this provides evidence that having file extensions that are related to the name of the application itself can help users with file management. The percentage correctly identified for Javascript (.js) was 87.7%, Structured Query Language (.sql) was 69.6%, Portable Network Graphic (.png) was 79.0% and Comma Separated Value (.csv) was 68.1%. These success rates have less to do with specific knowledge about the application and the file extension and more to do with making it easy for users to intuitively link file extensions to the applications they serve. Although it's easy to find out what application an extension serves via a quick browser query, typically what file management requires is the other way around. You know what application you used when searching for a file and knowing what the extension is, makes it easier to find. If that file extension is tied to the application name, it is even easier to remember.

Clearly, a comprehensive understanding amongst students of file extensions does not currently exist. And, the lack of growth in that knowledge over time is also apparent. Additionally, a student's personal device and corresponding OS doesn't have a significant effect on file extension knowledge when the academic environment requires every student to work in the tested environment. From this, we would recommend some remedial education for students that can be reinforced. Of course, file management can be taught as part of a course but some sort of repository of that knowledge should be made available to all students. We might suggest some simple videos built around a shared section of the university learning management system.

RESEARCH LIMITATION AND FUTURE RESEARCH OPPORTUNITIES

Given the narrow focus of our research frame and data collection, the results yielded in this study cannot be attributed to college students in general. A more randomized procedure that captures students from a variety of disciplines from across campus and different schools is necessary to corroborate the findings of this study. Additionally, this work doesn't allow us to understand file management practices as a whole. It only represents one small piece of a holistic model. At some point, the entire model should be empirically tested in order to understand the depth and breadth of student knowledge, skills and abilities associated with file management.

This research represents the first foray into understanding the file management KSA's of college students. What is needed is a comprehensive understanding of these KSA's in terms of the complete file management model including each theme and the tools available to practice file management. For example, a study could look at the student's ability to find working files or to organize files and folders for classes or projects. There are plenty of opportunities to continue to mine the pieces of the file management model along with need to test and associate those parts as a complete practice.

Lastly, given the variety research focused on file management behavior and tools, research needs to relate the ability to practice file management and the KSA's associated with it to the observed behaviors and efficacy of the tools that enable its practice. Our review of the literature indicates that the file management education of users has not been considered as a factor impacting the FM behavior of users and their evaluation of the FM tools.

SUMMARY

In this study, we assessed the file extension knowledge of college students in a Windows OS. Six hypotheses were tested. The results are summarized in table six.

Table 6. Summary of Results

Hypothesis	**Result**
H1: Students whose personal computer is based on the Windows OS will not know more about how to manage file extensions in a Windows OS than students whose personal computer is based on the Mac OS.	Null hypothesis not rejected
H2: Upperclassmen (Juniors and Seniors) will know more about how to manage file extensions in a Windows OS than lowerclassmen (Freshman and Sophomores).	Not statistically supported
H3: A majority of students will recognize common extensions associated with software they use in their courses.	Mostly statistically supported
H4: Students whose personal computer is based on the Windows OS will not correctly identify more common file extensions than students whose personal computer is based on the Mac OS.	Some statistically significant differences found
H5: Upperclassmen will correctly identify more common file extensions than lowerclassmen.	Not statistically supported
H6: A minority of students will recognize uncommon extensions associated with applications they are unlikely to use on a regular basis.	Statistically supported

As educators, it is our goal to understand where knowledge and skills are lacking and to address those deficiencies in our students as best we can. While file management seems like a small part of the learning environment we strive to create, the practice of FM is something that will be done throughout a student's academic and work career.

Understanding file types and their associated applications is important to file management practices impacting storage, organization, retrieval and sharing. This study demonstrates the variability of know-how amongst students in terms of file extensions. We would posit that this variability exists throughout the KSA's needed to practice good file management and further research is need to understand the education needs of our students in this area. Additionally, students are not acquiring file extension knowledge as they proceed through their courses. The lack of growth seems to indicate that students don't understand the need for or usefulness of such knowledge.

REFERENCES

Agrawal, N., Bolosky, W.J., Douceur, J.R., & Lorch, J.R. (2007). A five-year study of file-system metadata. *ACM Transactions on Storage (TOS)*, 3(3), 9.

Barreau, D. (1995). Context as a factor in personal information management systems. *Journal of the American Society for Information Science*, 46(5), 327–339.

Blanc-Brude, T. & Scapin, D.L. (2007). What do people recall about their documents?: Implications for desktop search tools. *In Proceedings of the 12th International Conference on Intelligent User Interfaces* (pp. 102–111). New York: ACM

Dinneen, J. & Julien, C. (2019). What's in Peoples Digital File Collections. *82nd Annual Meeting of the Association for Information Science & Technology* | Melbourne, Australia | 19 – 23 October, 2019

Dinneen, J. & Julien, C. (2020). The Ubiquitous Digital File: A Review of File Management Research. *Journal of the Association for Information Science and Technology*, 71(1), E1-E32.

Hardof-Jaffe, S., Hershkovitz, A., Abu-Kishk, H., Bergman, O., & Nachmias, R. (2009b). Students' organization strategies of personal information space. *Journal of Digital Information*, 10(5), 1–17.

Henderson, S. (2011). Document duplication: How users (struggle to) manage file copies and versions. *Proceedings of the American Society for Information Science and Technology*, 48(1), 1–10.

Henderson, S. & Srinivasan, A. (2011). Filing, piling & structuring: Strategies for personal document management. *In 44th Hawaii International Conference on System Sciences* (HICSS; pp. 1–10). San Diego, CA:IEEE.

Holmes, A, Illowsky, B. & Dean, S. (2017). *Introductory Business Statistics*. Houston, Texas: OpenStax.

Jensen, C., Lonsdale, H., Wynn, E., Cao, J., Slater, M., & Dietterich, T.G. (2010). The life and times of files and information: a study of desktop provenance. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 767–776). New York: ACM.