

Heather Chapman. The File Naming Habits of Personal Computer Users. A Master's paper for the M. S. in I. S. degree. March, 1999. 85 pages. Advisor: Stephanie W. Haas

This study is an extension of research by J.M. Carroll (1982) on the naming of personal computer files. Eleven one-on-one interviews were conducted with both veteran and novice Personal Computer users to gain insights into the filenaming habits of individuals and the reasons for each of these naming behaviors.

Randomly-selected filenames were examined both in isolation and within the context of other filenames, for morphological, lexical and semantic attributes and patterns; and evidence of the participants' purposes in employing each technique was obtained from the interview transcripts.

The range of different filenaming strategies observed in this sample has implications for both the design of future measures of filenaming behavior, and of improved file management systems.

Headings:

- Computer File Names
- Onomasiology - Reference (Linguistics)
- Information Storage and Retrieval Systems -- File Management
(Computer Science)
- Windows (Computer Programs)

THE FILE NAMING HABITS
OF PERSONAL COMPUTER USERS

by
Heather P. Chapman

A Master's paper submitted to the faculty of the
School of Information and Library Science of
the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

March, 1999

Approved by:

Advisor

Table of Contents

Introduction: The File Naming Habits of Personal Computer Users	3
Naming For the Recall and Retrieval of Information in Electronic Form	4
Method	14
<u>Study Participants</u>	14
<u>Possible Training Received by Participants</u>	16
<u>The Interviews</u>	18
Justification of Method	22
<u>Previous Work</u>	22
Results	24
<u>Collection and Analysis of the Data</u>	24
<u>File Types</u>	27
<u>Filename Lengths</u>	29
<u>Classification of Patterns in the Data</u>	30
<u>Categories on Character Patterns in Independently-Considered Filenames</u>	32
All capitals.	33
All lowercase.	34
Capitals and lowercase hybrid.	35
Spaces, etc.	37
Packing.	38
Blending.	39
Letter doubling.	40
<u>Categories on Lexical Conventions in Independently-Considered Filenames</u>	41
Full words.	43
Truncation.	45

Standard contraction.	46
Invented contraction.	47
Acronyms or initials.	47
<u>Categories on Semantic Properties in Independently-Considered Filenames</u>	48
Subject of contents and function of contents.	49
Dates and proper names.	53
<u>Categories on Properties in Filenames Considered In Context</u>	55
Isomorphism.	56
File type or extension.	58
Character or substring serial position.	58
Index expression.	59
Semantic contraries and semantic relates.	61
<u>Parts of Speech</u>	63
Issues Surrounding Filenaming	66
<u>When to Name</u>	66
<u>Renaming</u>	67
<u>Retrieval</u>.....	68
Summary of Observations	70
<u>Profiles of the Typical Filename</u>	71
Concluding Remarks	74
<u>Future Research</u>	74
<u>Implications for Basic Training in Computers</u>	76
Notes.....	77
References	79
Appendix A.....	83

Introduction: The File Naming Habits of Personal Computer Users

“The first business of real names is to successfully refer” says John M. Carroll (1985, p.97). Possibly the kind of name most tested in this regard is that which must successfully refer to something unseen and intangible, with limited observable properties - something like a computer file, which is when you get down to it, nothing more than a flash of static electricity on a chip. At least, without an interpretable, eye-catching, and memorable filename, a computer file meant for human consumption can be no more than that. If you can't find it, you can't open it. Surely, for most people, the act of creating a name for a file must be at times a fairly sophisticated task, requiring some conscious thought and planning time. However, practically speaking, when the user is working within the constraints of many software applications – with the ever-lurking specter of lost data close at hand – the first business of file naming frequently becomes nothing beyond a short-term means of quickly saving one's work before it is lost to a hardware or software malfunction. How do individuals balance these two realities of working with files on a personal computer in their day-to-day work? Anyone who works with computers for any significant period of time generally accumulates a rather large collection of files, which must be in as much need of a coherent system of organization and labeling as any collection of works.

The present study was an attempt to expand on the work of others by exploring the many different approaches used by individuals to name and organize their personal computer files and to investigate the real-world reasons individuals have employed these strategies. In light of the past two decades' worth of changes to the PC interface and most operating systems' capabilities (e.g. long filenames, more flexible file searching, content scanning capabilities, increased capacity and decreased costs of internal and external storage objects, spatial cues – or implied spatial cues), investigation into how these may have influenced file naming strategies people employ and their effectiveness in file management appeared well overdue. Additionally, this study also was designed to broaden present understanding of computer file naming by contrasting the practices of experienced with those of the less experienced PC users and identify differences worthy of future investigation. Analysis of the practices presently used by these different groups of users may help us to identify new methods developed through the trial-and-error experiences of veterans while staying aware of the natural proclivities of novice users. Such insights could be a valuable resource for the design of better file management systems for all users.

Naming For the Recall and Retrieval of Information in Electronic Form

Within the narrowly defined domain of personal computer file management, Carroll (1982, 1985) identifies the basic criterion for a good name: that it serves a set of specific goals. These goals – obvious to those with any experience working with computers – include such basic functions as easy retrieval from one's memory, both for the purpose of locating the file and

triggering some understanding of its contents; distinguishing one file from another; and - perhaps to a lesser degree now, considering the prevalence of a point-and-click interface - level of difficulty to type while at the computer (Carroll, 1985, p. xv, & 1982, p. 406).

The past two decades of advances in software and hardware have resulted in countless changes impacting the domain of personal computer file maintenance. The most profound of these, which has revolutionary implications for the study of computer file naming, has been the introduction of the long filename for users of Microsoft products, whose software and operating systems are widely used. Along with Windows 95 came VFAT (Virtual File Allocation Table), which makes it possible for the majority of PC users to create filenames using a maximum of 255 characters.¹ In order to preserve compatibility with DOS and earlier versions of Windows, Windows 95 uses VFAT to store a shorter 8.3 version² of the name in addition to the long version originally assigned by the user. In this way VFAT is backwardly compatible with the original FAT (File Allocation Table). Windows automatically generates a unique short name by truncating the original filename to six characters and appending a tilde “~” and a number to this shortened root³ (Prosise, 1996). The short version is only visible to the operating system and when the file is opened using older software versions⁴; otherwise the long version is visible as the file’s name. Prior to this, the majority of the PC world labored under the limitations of the DOS filename, with an eight character root.

Since John Carroll's early explorations of computer programmers' strategies for coping with the 8.3 filename-length convention of the DOS world (Carroll, 1982, 23.a, 1985), there has been little, if any, scholarly attention paid to the computer file-naming habits of computer users in the literature. Even with the advent of long filename capabilities in many different new PC operating systems and applications, field studies on personal computer file naming appear to have been left fallow. This is not to say that the phenomenon of name generation within the computing milieu has been left wholly unexamined during this time. Indeed, with the development of many more English-like computer languages, psycholinguists and other students of human cognition, memory, and information processing behaviors have turned their attention to naming styles used by programmers for the variables, data structures and modules, etc. in their code (Teasley, 1994, 18.a) (Weidenback, 1986) and for the commands which are part of various software applications or operating system utilities (Katzenberg & Piela, 1993) (Cramer, 1990) (Zoepritz, 1986) (Carroll, 1985, chapt. 4, pp. 67-90). Each of these works includes discussions of study results which provide at least some support for the proposition that naming based on a consistent structure of rules results in vocabularies that are more easily recalled by users. Some of these observers also have gone on to collect data which suggest that heightened recall did positively impact the task performance of at least a subset of their studies' subjects (Teasley, 1994) (Lee, Foltz & Polson, 1994). To any individual attempting to revisit a file saved months or years before, the level of satisfaction he feels for the name under which it was saved is usually equal to the level of

ease with which the file was found on the hard drive. Revisiting one's files is usually the ultimate goal of naming the file in the first place, although with increasing prevalence of networked PCs, labeling one's files so that others may anticipate their contents may often be a parallel aim. Still, many PC users name a computer file for their own reference needs.

Mark Lansdale's 1990 examination of the psychological aspects of personal information management helps us to identify cognitive factors in naming which may affect the user's ability to successfully retrieve a file from his PC. Lansdale proposes that information retrieval within the realm of personal filing systems involves two distinct psychological processes: "recall-directed search," analogous to Herrmann's theoretical direct-access from long-term memory (cited in Lansdale), which is usually followed by "recognition-based scanning" (p. 42). Recall-directed search is defined by Lansdale as the use of memory to pinpoint the location of desired information as exactly as possible. When dealing with computer files, this would range in exactitude from complete recall of the pathname and filename to just *knowing* in which main directory or subdirectories to search. The recognition-based scanning process would then be used by the individual at the point where recall leaves off to interpret the meanings of the names in a group of files sufficient to recover a memory that connects one of them to the desired information. In short, Lansdale advocates the interpretation of observed behavioral strategies for working with any information storage and retrieval system as a "a trade-off" between recall and recognition. That is, the filename an individual creates for a particular document, and the directory and

subdirectories in which he chooses to save it, can be seen as his attempt to shift his dependence more towards either recall or to recognition, depending on which proportion of each he believes to be to his advantage.

Of course, the degree to which an individual using a computer system may rely on recognition-based scanning is much more compromised than with other more natural storage and retrieval situations, such as the management of physical documents in an office. In his detailed observations of the ways professional and clerical office workers organize information in their offices, Thomas Malone identifies several main characteristics about the two “most important units of desk organization” he calls “files” and “piles” (1983, p. 105). Malone defines files as “units where the elements ... are explicitly titled and arranged in some systematic order (e.g. alphabetical or chronological)” and piles as units wherein “the elements ... are not necessarily titled ... and not, in general, arranged in any particular order” - at least, not intentionally (pp. 105-106). Whereas Malone’s files do roughly correspond to the hierarchical single tree system of directories and subdirectories in a computer, there isn’t really any organizational entity on a computer which can perform the two distinct functions of “finding” and “reminding” as effectively as piles in an office. Using Malone’s insights, one can identify several ways in which the computer is still at a distinct disadvantage when compared to physical piles of papers for retrieving a particular document (or group of documents). When dealing with physical papers, one enjoys the benefits of three-dimensional space (with certain limitations presented by realities like gravity or traffic flow). This provides greater

flexibility in adjusting the location of a document, which is not yet possible with the average PC. The GUI on most PCs is still a two-dimensional image on a screen with severely limited real estate. The individual searching for a particular document stored in a pile will enjoy a greater opportunity to use what one might call spatial memory to recall its location. Perhaps this is one more reason to look forward to the age when Virtual Reality filing systems are commonplace. One of Malone's participants benefited from a system of piles in that he was free to arrange his documents into a pattern of piles which corresponded to the individual steps within a particular process, effectively mirroring his workflow with his organizational system (pp. 101 - 102). An attempt to retrieve a document about which little else is remembered except at which point in processing it was last seen might be successful if one were able to leaf through the appropriate pile. With a computer, one would probably need additional reference points to search for that item.

Another function of piles which is not as easily achieved with a computer is the ability to defer classification of a document. Because categorization of some items can often be both difficult and time consuming, an individual may need to put it off in order to focus his mental resources on the content of the item. If you put off categorizing a document on a computer (i. e., naming it and saving it in a directory), you risk losing it. Consequently much of the categorization of electronic documents is done hurriedly, without adequate planning, which often results in an inability to remember the filename and the location of the document at a later time.

Physical pieces of paper in a pile are also superior to their PC equivalents, document images, which are only accessible via file viewer applications, such as Windows 98's Quick View. A physical document may be recognizable by its distinctive letterhead or the paper's color and texture; it may be on multiple sheets, stapled or paperclipped to an envelope or a business card. A system of computer file attributes, which Malone suggests might serve as surrogates for these physical traits, includes visual cues such as frequency of display, graphical depiction of file sizes, special locations on the screen and a color code to indicate priority status (p. 110). As of now, such a system is not possible with current computer systems, at least not with the degree of automation that would be necessary so that maintaining it would not be yet another troublesome task to avoid.

Various file viewer applications, which allow the user to quickly scan the contents of a file without first having to launch the program in which it was created, have become available during the last few years (See Livingston & Straub, p. 192) (Joyce & Moon, p. 62). Even though such utilities make inspecting the contents of one's files slightly less time-consuming, it still forces a one-at-a-time approach which is cumbersome. Plus the user is faced with the additional complications of software compatibility issues ([Windows 95 Professional, 1998](#)). So the user is again left to depend primarily on his recollection of the filename and location, which is subject to the degree of effectiveness with which he has categorized and labeled his informational units (or files). In short, he's back to square one. Considering the comparative simplicity

and speed with which one can glance through a handful of papers to find a particular document, it is clear why many people still prefer working with paper copies when it comes to keeping track of certain information.

Indeed it is in the domain of personal computers where effective naming strategies are still of great practical importance for recollection and retrieval of one's files. A good indication of the universality of the difficulties individuals experience in keeping track of their electronic files is the plethora of articles on the subject of file maintenance which can be found in many magazines catering to beginner or experienced computer users (Dunn, 1998; Beckman, 1996; Bott, 1995; Mamis, 1994; Press, 1990; Roberts, 1992; Robinson, 1995; Tackett, 1995; Tinker, 1995). Some articles explore pre-Windows 95 document name management software, such as Name-It and Long Filenames (Mamis, 1994) (Goldsbrough, 1995), which essentially support long filenames in older versions of software applications. In addition, file naming issues have been touched on - to varying degrees - in many of the popular computer book titles aimed at beginner PC users. Many of the earlier publications tend to concentrate merely on the technical aspects of naming a file (Hewitt, 1994; Inter Trade, 1996; Rathbone, 1992 & 1993). However, in the more recent books on Windows 95 and 98, writers are beginning to supplement the basics with general admonitions that filenames be "descriptive" and "meaningful" (Rathbone, 1995 & 1998; Jamsa, 1995). Some have even gone so far as to detail specific naming conventions designed to facilitate easier organization of one's files (See Magid, 1996; Halverson & Young, 1997, p. 83; Davis & Crosby, 1998, p. 21). Such

practical tips offered to users range from the approach of using non-standard file extensions, prefixing filenames with consistent numbering schemes (Roberts, 1992, 11.a;) or “code letters” (Ellingson, 1995, 14.a) - either to identify content or affect the order in which the files are listed in the directory (Person & Rose, 1997, p. 82 & 104; 1995, p. 96) - to the more involved and formal practice of numbering rather than naming files, which requires referral to a separate directory of names and content descriptions when the user wishes to retrieve a file (Tinker, 1995, 9.a).

In the 8.3 and long filename worlds, the three-letter extension is conventionally used to identify the format of a file. Some file naming guides recommend using non-standard three-letter combinations to identify one’s own personal collection of formats, such as “manuals” identified by “.MAN” or using “.LTR” for letters, (Black & Motz, 1989; Minter, 1990). However, other Windows-oriented texts explicitly advise against tampering with the standard file extensions, such as “.DOC” and “.TXT” because self-generated extensions will not be automatically recognized by Windows, disabling its ability to map individual files with their corresponding applications when the user elects to open a file directly from a file listing (See Person & Rose, 1995, 1997; Press, 1990, p. 24)⁵ Even with the recent release of new PC products, which allow for a legal filename length from eight characters (plus the three-character extension) to a maximum of 255 characters⁶ without additional software, some of these same naming strategies are still recommended to users of the new operating systems (Beckman, 1996, 15.a). Some have pointed out (Bott, 1995, 13.a), that today’s

more sophisticated file management tools and file content information display capabilities of recently-released operating systems may have made an effective file naming strategy slightly less imperative for average PC users (Robinson, 1995, 10.a). The adoption of file naming conventions or separately installed software applications which create thumbnail images of file contents are not the only aids to finding personal files being explored in the literature on electronic information storage and retrieval. Other approaches, and variations on them, have also been proposed. These other avenues range from automated content indexing or summarizing, usually made possible by various standards for labeling content structure for indexing or summarizing, such as SGML (Press, 1994); to type or format recognition via “magic numbers” or identifying strings (see Hardy, 1996, p. 182). However, as David Beckman reminds us, the filename remains the best - and most “low-tech” means of facilitating more efficient retrieval of one’s work, which leaves the user less vulnerable to those mishaps or confusions which invariably lead to misplaced files (Beckman & Hirsch, 1996).

Continual advances in cheaper hardware with increased data storage capacity; improved file compression; and the recent years’ shift to networked PCs, which has revived the prevalence of shared files, all contribute to heightened difficulties for a computer user to find a particular file (Black & Motz, 1989, 6.a). These factors, when considered against the reality that a growing percentage of the every day information with which we all must work comes in electronic form, strengthen arguments for paying closer attention to personal computer file naming and organization issues.

Method

Study Participants

In order to learn more about how individuals name and organize their personal computer files, as well as the difficulties they experience when trying to determine the contents of a file they have named earlier, a series of eleven short and informal interviews were conducted with PC users. Volunteers for these interviews were sought from among both graduate and undergraduate students pursuing Information and Library Science degrees at a major University. The members of this population of students were of a wide range of ages and a variety of academic and professional backgrounds. Students with both high and lower levels of computer experience were sought for the study. An individual's enrollment in one of several specific courses designed for either lower or higher PC experience was used as an aid in accepting the desired mix of volunteers. Individuals enrolled in an Introduction to Computers course were considered "less experienced." Enrollment in a course which required the introduction course as a prerequisite suggested "experienced" computer users. These categorizations were later adjusted as more information was acquired about each person during the interview when he was asked to estimate the average number of computer files he named during a typical week of work and how that compared with his level of file-naming activity during any past periods. If an individual previously classified as "less experienced" due to the courses he was taking estimated a high enough number of file namings per week, he would be re-classified as "experienced".

The criterion of course enrollment proved to be an imperfect indicator of computer filename experience among study volunteers. The only male volunteer who was enrolled in the introductory computer course had more than 10 years' experience working with a Macintosh, only lacking experience with IBM-compatible machines. Consequently, there are no "less experienced" males included in this study. However, the three women volunteers enrolled in the same course fit comfortably in the "less experienced" category. Each of the three reported creating or naming fewer than ten computer files during a typical work week, with one person who never exceeded five. Conversely, another female volunteer was re-categorized as "less experienced," because of her comparatively low amount of file naming activity since completing an introductory computer course the previous semester. Additionally the less experienced participants indicated that their current averages were an increase over any past file naming activity. Indeed one of these students had never used a computer prior to coming to graduate school, relying instead on a typewriter.

Individuals considered "experienced," with the one exception described above, had already either completed or waived the Introduction to Computers course, and all reported naming at least ten files during a typical work week, either currently or in the past. The highest number of files named per week were estimates of 25 and 30. Both of these estimates were claimed by male volunteers, and the most experienced female volunteers reported highs of only ten or twelve per week. It is important to remember that these numbers are only the participants' estimates. A future study of filenaming would benefit from a

more concrete measure of filenaming activity levels. The final sample consisted of four “less experienced” females, three “experienced” females, and four “experienced” males, ranging in age from early twenties to their early forties.

TABLE 1.
Experience Levels Among Study Participants

Gender	Age	Number of Files Named Per Week	Past Weekly Filenaming Activity	Experience-Level Classification	Currently Enrolled ^a
Female	18 - 25	2 - 5	0	Less Experienced	Yes
Female	18 - 25	2 - 3	0 - 1	Less Experienced	No
Female	18 - 25	8 - 10	2	Less Experienced	Yes
Female	35+	8 - 10	0	Less Experienced	Yes
Male	26 - 34	30	10	Experienced	No
Male	18 - 25	4 - 5	15	Experienced	Yes
Male	35+	25	0	Experienced	No
Male	35+	3 - 30	3 - 10	Experienced	No
Female	26 - 34	5	10	Experienced	No
Female	26 - 34	0 - 1	0 - 10	Experienced	No
Female	18 - 25	2 - 10	2 - 10	Experienced	No

^a All participants not currently enrolled in an introductory computer course had either already taken it or were able to waive it.

Despite some initial difficulties in attracting participants, a review of Table 1 above shows that the eleven people interviewed for this study represented a fairly broad range of ages and computer experience. It was anticipated that a varied sample would lead to a richer variety of file naming strategies observed during the study.

Possible Training Received by Participants

Logic dictates that computer users do not develop their individual methods of file naming and organization in a vacuum, but receive instruction and examples from many different sources, such as course lectures, consultation with peers, or strategies they may have observed in the work of others. Contemporary written sources of instruction were readily available to the individuals included in this study. A wide variety of periodicals and textbooks on PCs are available through

the school's computer labs and the University Library system. A comprehensive survey of all these resources is not attempted here, nor were participants in this study asked to list outside influences on their chosen strategies for file naming, as this would fall outside the main focus of the study. However a brief review of at least a portion of the written materials generally available to these students is useful in characterizing the nature of instruction in file naming a novice computer user might encounter within required course materials and in popular computer texts. Syllabi for the introductory computer courses offered at the time of the study list two titles as required reading for instruction on file management, the Academic Technology & Networks online document entitled, "Windows 95/98: File Management," and the O'Hara & Vega text, Discovering Office 97. Of course, any expectation that most students, let alone this study's participants, read and were influenced by the file naming practices depicted in these materials should be tempered with the realization that the mere presence of information is no guarantee that it has been used.

However, for those who do consult the texts, the O'Hara book, and ATN's online documentation offer some practical instruction on naming files and directories. ATN's online material provides a few simple examples of names for files and subdirectories. For situations where files are shared, it does recommend developing an agreed-upon naming convention based on easily remembered rules, which can accommodate any growth in the file collection. However, no specific naming schemes are offered. The possibility of long filenames is mentioned, but all the naming examples supplied are single-words

with eight or fewer characters and three-letter extensions. The materials by O'Hara go into much greater detail by providing many more contrasting examples of files labeled via several different naming strategies. Neither of these sources espouse any particular approach to naming files. They instead tend to leave the specifics up to the reader, stating that any good file naming practice must be compatible with the individuals using it and the context of the work.

Many of the popular computer books described in the preceding sections of this paper were also available at the computer lab frequented by the participants in this study. Although the lab's collection tended to include the most helpful and detailed of the computer texts covered in the preceding literature review (Person & Rose, 1995 & 1997; Halverson & Young; Davis & Crosby, 1998; Weverka, 1996; Boyce, 1998), most instructional materials for PC users appear to treat explicit instructions on file naming strategies more as an afterthought or bonus subject than as a basic skill for effective computing.

The Interviews

Each of the 40-minute interviews were one-on-one and were all audio-taped. The amount of observational data lost by not recording a videotape of the interviews was judged to be less important than avoiding the level of self-consciousness that some participants might have felt at being videotaped. The audio tape was readily ignored by the participants, and its presence enabled the interviewer to better concentrate on consistently and clearly framing her questions. It was also a good way to provide a much more detailed account of the participants' responses than would be possible with post-interview notes,

since many of the interviews were scheduled back-to-back in order to accommodate participants' schedules and complete all eleven sessions within a single week.

The interviews took place at a school computer lab, with the participant sitting at a Dell Optiplex Pentium II 400Mhz PC running on the Windows 98 operating system. The interviewer sat to one side, affording a clear view of both the participant and the computer screen. Prior to each interview, all viewer and data import utilities were disabled and all settings on the computer changed so that only the filename and extension, Windows "object" icon, file size and type, latest "modified" date and time for each file in a directory would be visible. Filenames were viewed using the Windows 98 "My Computer" file manager and all listings were set to appear in a single window, sufficiently enlarged so that the entire pathname could be viewed in the title bar. (See Appendix A for a sample screenshot.) At the beginning of each interview, the order of all listed files and subdirectories was alphabetical by filename. Participants were free to select other ordering options (i.e. by size, type, or modified date) during the interview.

Interview questions focused on the processes used by individuals to name and rename their files and directories and the reasons behind them. Although the same basic set of questions (see Figure 1) was used for each participant, the different responses they elicited, as well as the informal, semi-structured nature of the sessions, guaranteed some variation in how the questions were presented and what follow-up questions were asked. The interviewer began each session by asking the participant to estimate the average number of files he named

during a typical week of work. He was then asked whether his estimate was an increase or decrease compared to the average amount he named during a previous situation. (What is meant by a “previous situation” is perhaps best communicated with the following example: A student working a full-time programming job over the summer sees a change in his file naming activity level when he switches to part-time work and begins a new weekly routine of attending classes.)

At this point, each participant was asked to find his personal directory and display the files and subdirectories (or subfolders) in that directory on the computer screen. The majority of those participating in this study kept a personal directory on a common drive, a storage space networked to all the computers in the lab and accessible to all lab users. Others brought in a Zip™ disk or a series of 3.5-inch diskettes. The interviewer used a random number table to select four files from the directory⁷ and asked the participant to describe how he created the name, detailing the reasons or processes behind his choice of structure and composition for each. Files of a non-text or non-graphic type, which were not recognized by the operating system, such as executables or custom file types created by the participant, were excluded from consideration. Another reason for exclusion would be if the participant indicated that the selected file was actually named by someone other than himself. In these situations, another file was chosen in its place. This occurred on four occasions during three of the interviews.

As the participant answered questions about each of the selected files, the interviewer took screen shots of the file listing to augment the interview recording and notes. Towards the end of the interview, the participant was asked to move to where he could no longer view the computer screen. The interviewer then randomly selected a file and printed a hard copy of its contents with a laser printer.⁸ The printout was given to the participant, who was instructed to find the corresponding file⁹. Each participant was free to refer to the file listings in his directories, but not to open any of the files. The participant was asked to make three attempts to name the filename, but whether or not each attempt was correct was not revealed until after all three attempts had been made. This constraint may have resulted in a slight difference in the search strategy used by the respondent than might have been the case in a real world situation, where the person would be free to inspect the contents of one file after another until the correct one was found. However, it was felt that this restriction was necessary in order to encourage each participant to concentrate on the role filenames played in the process of finding the file with the desired contents.

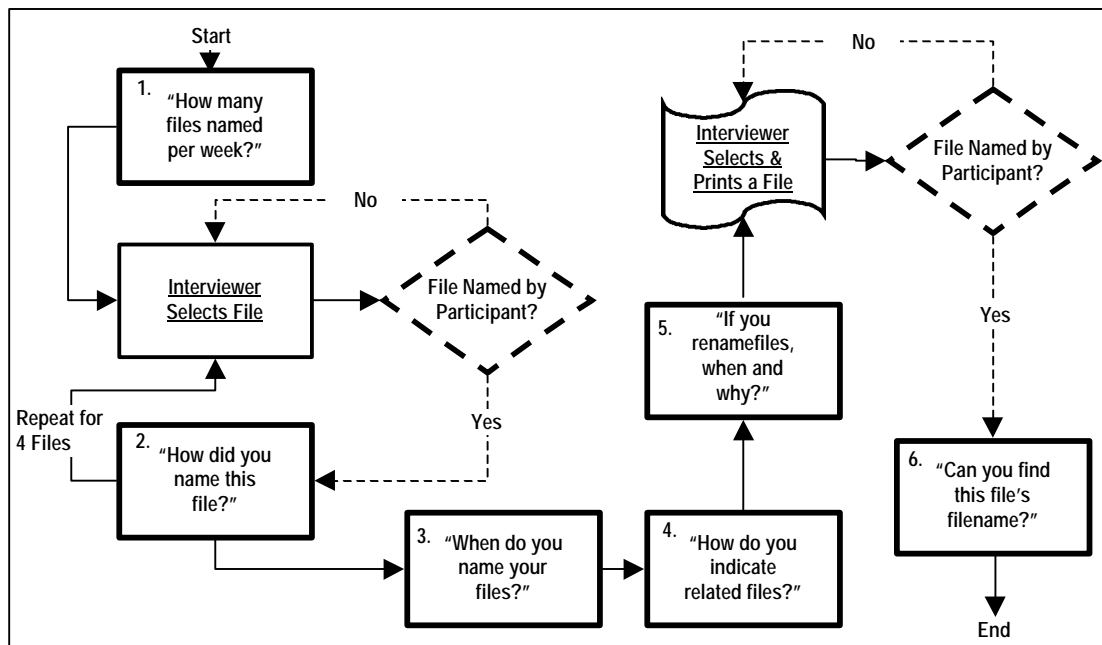


Figure 1. Summary of the basic question and task sequence for each interview. Different participant's responses caused variations to this sequence in some interviews.

Justification of Method

Previous Work

At present, the only study done within the domain of computer file naming appears to be the work of John M. Carroll, who looked at over twenty-five hundred filenames produced among twenty-two professional computer users at a scientific research center. (1982)

In this study, individuals were asked to compile lists of filenames from their personal directories, annotate each filename with a short description, and send it in electronic form to the researcher. Each participant prepared this list at his own computer without time constraints or supervision. Individuals were asked not to examine the contents of the files when writing their descriptions and to indicate which files they could not remember. A total of 2507 filenames were analyzed for repeating patterns of structural configurations, called "rule-schemes." Rule-

schemes are defined by Carroll as “systematic ways in which the composition of lexical items may suggest properties [such as the content] of their referents” (1981). Additionally, “inferred filenames” were reconstructed from the filenames using the written descriptions provided by the file’s creator and inferences about the abbreviation strategies used to make the name conform to the eight-character filename length limitation imposed by the MS-DOS operating system used at that time. These inferred filenames were then analyzed for grammatical and syntactic patterns as well.

The decision to take a qualitative, rather than quantitative approach to studying filenaming by PC users hinged on an observation stated repeatedly in much of Carroll’s work. Naming is goal directed. A quick perusal of the results from the many other studies on naming by Carroll (1985) and others (Collantes, 1995; Cramer, 1990; Katzenberg, 1993; Hodge, 1973; Teasley, 1994; Weidenback, 1986) suggest that the larger context within which the act of naming is performed may have a significant impact on what naming strategies will be most prevalent in the results. In other words, it’s entirely likely that a study wherein people named files in a controlled setting, removed from the real constraints and overriding concerns that are usually present when an individual is working on a computer, will fail to catch the many different ways people cope with the difficulties of managing hundreds of files. Indeed, during the current study, it was common for a participant to apologize for the state of his computer files, then describe how they would be if there had been sufficient time to go back and rename and re-arrange them prior to the interview. However, it is good that

none of the individuals interviewed found the time to do this, as the whole point of the study was to find out what *is* done - not what should be.

Results

Collection and Analysis of the Data

Once the collection of data was underway, it quickly became apparent that limiting the interviews to discussions about only the five randomly-selected filenames¹⁰ would overly restrict the participants' abilities to explain their naming strategies and the reasons behind them. Each participant brought anywhere from one to fourteen additional filenames into the discussion on his own, often in order to illustrate the purposes served by different aspects of the randomly-selected files' names. To ignore these additional data would have seriously blinded this study to the very real impacts on file naming that other files with related content (or which merely occupy the same directory) can effect. However, because of the differences in how these files were selected, their presence in any numerical surveys of the filenames in the sample would probably result in an artificial bias in favor of those naming approaches employed by participants who contributed the larger number of additional files during their interviews. Consequently, these additional fifty-four files were not combined with the originally-chosen fifty-five filenames for any of the quantitative descriptions in this study. However, their presence was taken into consideration during the inductive process of constructing a framework of categories for classifying the many different structural and conceptual patterns employed by the participants in creating their filenames. Sources used to illustrate the detailed discussions of

each category of naming aspects also included selections from the participant-volunteered filenames, in addition to the original randomly-selected names.

As stated earlier, the focus of this study was not only the structural and conceptual patterns evident in the filenames themselves, in conjunction with their creators' comments, but also the individual's reasons for choosing certain naming strategies and attributes. The more interactive method of data collection, via face-to-face interviews, was employed instead of the survey-like approach utilized in Carroll's study, so as to obtain more direct evidence of the reasoning behind the artifacts. Apart from a few follow-up interviews he made for clarification purposes, Carroll's explorations into why an individual named a file in a certain way was limited to speculation or what he was able to deduce from the name itself and its exegesis, as written by the file's creator.¹¹ During the interviews for this study, evidence of the specific motivations behind the each participant's choices in naming a file was also collected. However, one must assume that there are other reasons underlying the particular naming practices that were not verbalized during these interviews, but which may exist in the experiences of personal computer users. Due to the dynamic nature of the organizational system being studied here, the option of follow-up interviews was not exercised. Without taking a data dump or a complete collection of screen shots at the time of each interview, there was no practical way to insure that a participant's file directory contents and organization would remain unchanged through multiple interviews. This would be either as a result of the individual's

normal working activities or because they made purposeful changes inspired by issues discussed during the initial interview.

Additionally, in order to limit the number of preconceptions the interviewer could bring into both the interviews and her analysis of the data,¹² close examination of existing analytical frameworks was avoided as much as possible until after the collection and initial analysis of the data was completed. This is in accordance with the grounded approach to analysis as described by Taylor & Bogdan which describes the practice of some to distance themselves from existing research prior to completing an independent examination of their data (1984, p. 126 - 129). These measures may have caused the interviewer to neglect to ask some questions related to patterns present in the filenames that were only recognized after the interviews were finished. Consequently, some gaps were left in the lists of motivations behind those filenames evidencing characteristics not noted during the interviews because the questions were not asked.

For some of these name characteristics, such as the use of an index expression¹³ or of a common substring in multiple filenames, there was a temptation to assume the filenamer's reasons for adopting the approach. However, this kind of interpretation was avoided as much as possible, and a more conservative approach was taken by not assigning motivations to the use of specific naming patterns unless there was direct evidence of this, such as the individual's own words or, in some cases, gestures (e.g. shrugging the shoulders). On many occasions a participant would declare that he did not have

a reason for his employment of a particular attribute when composing a filename, suggesting that a certain portion of the motivations behind how individuals name their files may be subconscious or an inadvertent response to certain long-forgotten circumstances (e.g. lack of time, other filenames subsequently deleted, etc.). So even the person who is most likely to know, the filename's creator, can only speculate as to his reasons for his own actions.

File Types

Because of file types not recognized by the Windows 98 operating system, or whose corresponding applications were not available at the lab where these interviews were conducted, the range of file types encountered during this study is less varied than those observed by Carroll. The categories of programs, procedures and "miscellaneous"¹⁴ are not found here (See Carroll 1982, p.408-9). As shown in Table 2 below, of the fifty-five randomly-chosen filenames, fifty-one fell neatly into a broad category of text files, while the remaining four were image files.

TABLE 2.
File Extensions Encountered in Study

File Types		Number of Randomly-Selected Files	Portion of All Randomly-Selected Files (%)	Number of Individual Contributors
Text Files	total	51	92.7	11
.doc - "document"		34	61.8	11
.rtf - "rich text format"		4	7.3	4
.txt - "text"		2	3.6	2
.ppt - "power point"		2	3.6	1
.flo - "ABC flowchart"		2	3.6	1
.html - "hypertext mark-up language"		4	7.3	3
No extension (text file)		3	5.5	1
Image Files	total	4	7.3	3
.jpg - "joint photographic experts group"		3	5.5	3
.tif - "tagged image file format"		1	1.8	1

Note. Counts are of randomly-selected files only.

As is shown in the above table, the majority of text files were Microsoft Word documents. Every one of the participants' set of randomly-selected files included at least one name sporting a .doc extension. Among the fifty-four additional participant-volunteered files, a few additional extensions appeared, including .wpd (Word Perfect document), .gif (graphics interchange format), .xls (denoting an Excel spreadsheet) and some others designating files created using common statistical packages. But these too did not present any differences judged sufficient to require they be considered separately in terms of analysis. As it turned out, the only randomly-selected files encountered during the course of an interview that were rejected were discarded because the participant was not the individual who named them.

Filename Lengths

According to the percentages displayed in Table 3, the most prominent change to the PC filenaming domain since Carroll's work during the 1980's, long filenames, appeared to have received cautious acceptance by the individuals in this study's sample.

TABLE 3.
Filenames In Excess of Eight Characters

Participant Groups (Member totals)	Proportion of Total Filenames by Each Group (%) ^a	Number of Individuals Using Long Filenames ^b
All Participants (11)	43.6	10
Less Experienced Group (4)	55.0	4
Experienced Group (7)	37.1	6

Note. Counts are of randomly-selected files only. N=55.

^a Percentages are based on the total filenames contributed by all the participants in this experience-level. For the Less Experienced group, N=20; for the Experienced group, N=35; for all participants, N=55.

^b These counts are of the number of individuals within each grouping (All participants, Less Experienced, and Experienced) who had at least one long filename among their five randomly-selected files.

As shown in Table 4, the average filename length was above eight, both overall and for each experience level. However, the most prevalent filename length among the twenty files contributed by the Less Experienced and the thirty-five named by Experienced participants, and for all files overall, was still eight characters or fewer. This suggests that these participants perceived continued advantages to keeping filenames short which outweighed any benefits from the increase in naming flexibility made possible through use of longer filenames.

TABLE 4.
Filename Lengths

Experience Level Groups	Average Filename Length	Median Filename Length	Mode
Both Levels	11.3	8	8
Less Experienced	13.8	8.5	4
Experienced	9.8	8	5

Note. Counts are of randomly-selected files only. For the Less Experienced group, N=20; for the Experienced group, N=35; for all participants, N=55.

Classification of Patterns in the Data

The classification scheme used to categorize filenames according to their structural (or morphological) and conceptual characteristics was drawn mainly from patterns observed in the data, as well as some evidence gleaned from the researcher's past experiences with personal computer users. Refinements to the scheme, and the ideas for some additional categories and terms, were adopted from Carroll's system of analysis (1982). The final product was twenty-two categories of filename characteristics, aggregated into four broad sets (see Table 5 below). These four sets were based on two different ways of considering each filename, independently or within the context of surrounding files, combined with three levels of abstraction at which to analyze them. At the first level of analysis one considers the morphological properties of the individual letters, numbers and symbols which are contained in a filename. For example, considering the filename "FinalRptGWcover.doc" at this level of analysis, one would focus on features such as the combined use of both capital and lowercase letters and the lack of spaces or symbols. In analyzing the same filename at the second level, at which the lexical form of the substrings is considered, the scope would broaden to consider certain characteristics of the substrings within the filename. At this level of analysis, the substrings "Final," "Rpt," "GW," and "cover." are each classified as either a full word or an example of an abbreviation type. The third and final level of analysis (and the most abstract) is concerned with the semantic properties of the inferred words¹⁵ represented by the substrings in a filename. At

this point, one would be focusing on the meanings of “Final,” “Rpt,” “GW,” and “cover,” as well as the purpose served by each the filename or within in the context of surrounding filenames. Because of these multiple layers of observation, it was practically a given that a single filename would exhibit characteristics which placed it in multiple categories.

TABLE 5.
The Four Main Sets of Categories

Category Set	Frame of Reference	Unit of Analysis within Filename	Level of Analysis	Number of Member Categories
First Set	Out of Context	Character	Morphological	7
Second Set	Out of Context	Substrings	Lexical	5
Third Set	Out of Context	Substrings	Semantic	4
Fourth Set	In Context	Substring Sets	Semantic	6

The first three of the four category sets encompass sixteen of the twenty-two categories to which it is possible to assign (or exclude) a filename which is considered in isolation from any other filenames. For example, it is possible to determine whether lowercase or capital letters - or a combination of the two - have been used in a filename without referring to the other filenames in the directory listing. The fourth group of categories distinguish themselves from the rest mainly by virtue of the fact that their patterns are only evident in a filename when it is considered along-side its companion files. These approaches are analogous to the “rule-like strategies for imposing redundancy” which Carroll calls “rule-schemes” (1982, p. 409). In other words, this group of filenaming strategies, like Carroll’s rule-schemes, are only visible in the context of the other filenames in the file listing.

It is important to note that all portions of the analysis were conducted under the same conditions, namely that the scoring was conducted by a single

person. For this reason the counts offered in the following sections were subject to some error, even though scoring for each category was characterized by repeated passes through the raw data and numerous recounts. At the same time, special care was taken to define each category rigorously, removing judgement calls to the greatest possible extent. Additionally, because of this one-manned methodology, it is possible that some filenames which would properly belong to a particular category may simply have been overlooked.

The following descriptions of the data, and their analysis according to the twenty-two categories of filenaming approaches, are augmented by four tables, one for each Category Set. Each table presents a statistical summary of the category assignments for the analyzed filenames. The figures shown include the "Proportion of Filenames Overall" for each category, which is the percentage of all the randomly-selected filenames created by both experience-level groups. Also displayed are the percentages of filenames for each category, broken down by experience-level group. Additionally, the numbers of individuals from each experience-level group who contributed filenames which fit into a category are listed, to better reflect the popularity of particular naming practice within the two groups.

Categories on Character Patterns in Independently-Considered Filenames

The first of the three category sets used to classify independently considered filenames contains those categories concerned with the use of characters, including letters, symbols, punctuation and spaces, in the filenames.

TABLE 6.
First Category Set

Category	Proportion of FileNames Overall (%)	Proportion of Total L.E. ^a FileNames (%) ^b	Number of L.E. ^a Contributors	Proportion of Total Exp. ^a FileNames (%) ^b	Number of Exp. ^a Contributors
All Capitals	9.0	5.0	1	14.3	3
All Lowercase	38.2	30.0	4	42.9	6
Capitals & Lowercase	40.0	60.0	4	28.6	3
Spaces, etc.	27.3	35.0	2	20.0	6
Packing	36.4	20.0	3	45.7	3
Blending	1.8	0.0	0	2.9	1
Letter Doubling	3.6	0.0	0	5.7	2

^a Randomly-selected filenames created by the “Less Experienced” and “Experienced” participants.

^b Percentages are based on the total filenames contributed by all the participants in this experience-level. For the Less Experienced group, N=20; for the Experienced group, N=35.

It should be noted that at least four of the strategies encompassed by the categories listed in Table 6 above (i. e., the three variations on use of capitals and/or lowercase letters, and the use of spaces and certain symbols.) were not options for personal computer users at the time Carroll conducted his study of filenames.

All capitals.

The first category in this group, labeled “All Capitals” refers to filenames in which every letter is in the uppercase.¹⁶ Individuals employing the All Capitals approach to filenameing constituted a minority of the sample. Only one of the less experienced and four of the experienced participants had any filenames belonging to this category. Among these individuals, the use of all capital letters was not extensive. Only six of the fifty-five randomly-selected filenames employed it.

When asked why only capital letters were used in these instances, most participants attributed it to an inadvertent keystroke activating the Caps Lock feature, together with their lack of desire to devote time and effort to renaming a

low-priority file. One individual did purposefully elect to capitalize all the letters in her filename in order to differentiate between it and some otherwise identically-named version of it so that she could delete one of the versions without accidentally deleting the file she wished to keep. Other individuals appeared to capitalize their filenames when an acronym was used for a filename, as if applying a convention of written English. However, there was no explicit statement of this from any of these participants, so this reason must be considered mere speculation. Another individual claimed that the reason for her use of all uppercase letters in a filename was due to a default setting on the computer she was using at the time.

All lowercase.

Compared to the All Capitals practice, two approaches, the use of all lowercase letters or a combination of both lowercase and capitals, were much more commonly used by the participants in this study. With the All Lowercase category, everyone, save one member of the experienced participants' group, had at least one filename with only lowercase letters from among the five randomly-selected files. Overall, more than a third (38.2%) of these fifty-five filenames were classified as such, with a slightly larger proportion of the filenames made by the experienced participants evidencing this approach (43%) than among the less experienced persons' files (30%).

Once again, the individuals' stated reasons for this choice in naming tended towards a combination of habit and reducing the number of necessary keystrokes. As one person expressed this, "It [capitalization] is one more shift to

do.” Another individual expressed this even more explicitly: “...it may not even be that thought out. [...] It could be because I don’t want to do the keystroke; or it could be because I just don’t do it. I don’t really know if there’s really a reason.” Still another person avoided capitalizing to lessen the risk of accidentally hitting the Caps Lock key instead of the Shift key. Another mentioned the influence of other users on her decision to use only lowercase for files she expected to work with in different operating systems: “...I looked at someone else’s and they did it that way and they didn’t have any problems... I had been having a lot of problems, so I just said, ‘To be safe, I’ll copy exactly what they have done.’”

Capitals and lowercase hybrid.

The proportion of filenames overall with a mixture of both lower and uppercase letters was about the same as that of filenames with no capital letters at all. However, this practice was twice as frequently encountered in the less experienced participants’ files than was the case among the more experienced.

There was a range of reasons given by the participants’ who produced the filenames in this category. Three of these seven individuals (experienced as well as less experienced) indicated that this aspect of the name was actually suggested by the application or operating system which they were using. “...it [Word] almost invariably tends to capitalize it [the initial letter of a word], and it’s a pain in the neck to try to un-capitalize it, so I just let it go.” For another student, some of these filenames were actually attempts to use all uppercase letters, again frustrated by the software:

I try to do a capital and it won’t take it [...] it gets mixed up, and I think “Oh, just forget it!” [...] I wanted ... one to be all caps, and it

didn't take it the first time or two that I tried it, so I just - "Forget it."
There probably is a way to do it, but - [*shrugs shoulders*].

Beyond following the path of least resistance, some participants actively chose the combination of upper and lowercase letters to mark a separation between the substrings of a multi-word filename, without resorting to inserted spaces or symbols, such as hyphens or underscores. "...I tend to use the capitalization to give me a visual cue that there's a word break there." Another closely related reason was to attain a visual clarity for easier interpretation of the words inferred by each substring. One individual discussed his efforts to clarify the appearance of his file "URLnotes.doc": "...the word 'notes' is in lowercase because there are three uppercase letters just before it, and if I put 'notes' in uppercase, it would look like one big word." Still others ascribed their use of both cases in a filename as an aesthetics- or habit-driven allegiance to the conventions of written English. One of the experienced participants, who was the only former Macintosh user, offered this explanation: "I think I capitalize all my file[name]s as if I was naming a title at the top of the page..." In others, it was a combination of the first two reasons of separation and clarity together with a nod at proper written English:

[FinalRptGWcover.doc] ...I wanted to signify a new word, so I used the capital 'r.' 'g-w' is a proper name [Glaxo-Wellcome], so I used 'GW,' and with 'cover,' I wanted to distinguish that from the 'w' - although it [cover] is a new word too...

Another individual also attributed her habit of capitalizing only the initial letter of the first substring in a multi-word filename to her experience typing book titles

onto the cards for an old-fashioned card catalog, a domain where this practice is the proper form.

Spaces, etc.

The “Spaces, etc.” category encompasses two slightly different approaches to filenaming: the inclusion of spaces versus the insertion of symbols, some of which in a sense mimic a space. (e.g. the underscore or hyphen). Four individuals out of the eleven interviewed (2 experienced and 2 less experienced) used actual spaces in one or more of their filenames. The two less experienced individuals indicated that the spaces were suggested by the applications they were using, and they had accepted these suggested filenames with few changes, in lieu of typing in another whole filename themselves.¹⁷ The two individuals from the experienced group who kept spaces in some of their filenames, also listed the separation of substrings and visual clarity as reasons for their use of spaces; although one individual expressed some trepidation about their effect in different software systems:

[3.3 Sources.ppt] It's typical because Microsoft products allow you to do that, and it's natural; but I ran into a problem recently when I put something into Unix and it can't [sic] read the space. So I've gotten into the habit now of either compressing them all with no spaces, or putting a line marker in there...like an underscore.

One member of the less experienced group also included punctuation as well as spaces in her filenames, inserting “periods” in what appeared to be habitual deference to the conventions of written English: “[ENCYCLOPEDIA EVAL. KINGFISHER CHILDRENS.doc] ...it's...just an abbreviated form of evaluation. I never thought about that.”

The remaining three categories included in the first category group are approaches that were technically available options to the filename creators observed by Carroll. However, his analysis did not include them as separate categories. This may well be due to the changes to the filenaming domain that have occurred since that time, epitomized by PC users' present ability to create longer filenames with a much larger array of legal characters. This makes a category like "Packing" more of an option to PC users than the near-requirement it was for the computer users at the time of Carroll's study.

Packing.

The phenomenon of "packing," which is defined here as the practice of placing two or more substrings within a filename with no intervening blank space or symbol, can be considered the logical opposite of the "Spaces, etc." There were only three participants of the eleven volunteers who did not have at least one file among their randomly-selected five in which packing was used. These three were all members of the less experienced group. The twenty percent of less experienced user's filenames with packing were all contributed by one person, whereas the near 46% of experienced users' filenames in which packing was used came from all but one of the seven in that group.

For the majority of individuals from both groups, the sole reason given for this practice was the need to accommodate older software systems which did not recognize spaces or certain symbols.

[AltaVistaAids.ppt] ...I ran into a problem the first time I created a document in Word, saved it in html and tried to put it up on the Web.In the process of moving it to the Unix server, document

names with spaces in it [sic] got lost, so I just started to compress them back together...

Only one member of the experienced group gave an additional reason for her use of the packing approach to filenaming. She expressed annoyance at others' use of spaces and underscores in their filenames, preferring to save keystrokes and rely on her ability to figure out the inferred words of a filename herself:

...a lot of our stuff had underscores, and it just annoyed me. It's just an extra keystroke that isn't necessary. I mean, like you can't tell what it is unless you get the underscore in there?I can figure that out all on my own!

Blending.

The term "blending" is used by Carroll to describe the contraction of two words into a single term, which he illustrates by using the example of "breakfast" combining with "lunch" to make "brunch." However, "Blending" for the purposes of this study, refers to a less dramatic merging of two words or substrings via the sharing of a letter that they both have in common. Carroll considers his version of "blending" to be a type of abbreviation. Granted, even in its current definition, Blending does involve the technical removal of a character from a full word, which is a common characteristic of all the categories featuring different types of abbreviation. However, as a participant pronounced a filename that exhibited such blending, the manner in which he read it constituted indirect evidence that he was using that single letter shared by the two adjacent substrings for *both* words. In a sense, the letter was being *read twice*. From this perspective, there were no letters missing from either substring because of the blending procedure, only the spatial *width* of one character. However, this is not to suggest that the

participants spent any time pondering these subtleties. Only one file out of the one-hundred and nine encountered during this study could be placed in the Blending category. This filename, “210batstart.doc,” was described by its creator as a selection of short story “beginnings” about “bats.” Unfortunately, her reasons for having the words “bats” and “start” share the letter ‘s’ were not clear. The reasons one might suppose would be behind such an arrangement could include a desire to limit the length of the filename, an aesthetic objection to identical characters being placed adjacent to each other, or an uncorrected typographical error.

Letter doubling.

Filenames wherein matching characters, each from different substrings or complete words, are left adjacent to each other with no intervening spaces or symbols, are all classified under the category of “Letter Doubling.” Letter Doubling and Blending pair up as logical opposites, both within the space constraints present in filenames which fall under the “Packing” category. A filename on which the approach of Letter Doubling has been used will contain words or substrings for which the last letter of the first substring will match and fall immediately next to the first letter of the second substring, forming combinations like “starttime.”

Only three filenames of the randomly-selected fifty-five fit into the Letter Doubling approach category, and each of these were contributed by a different member of the experienced group. One of these filenames, “228--3.doc” differed from the other two in that the character doubled was the hyphen symbol. This

was in fact the result of a slip of the finger while typing. Another, “usefullinks.doc,” came about for no particular reason, its creator could remember, save an allegiance to correct spelling together with an earlier-expressed desire to avoid problem-causing spaces. The third “Fammhome.doc” contained material for the “Family Medical Homepage,” but no specific reason for the “mm” portion of the filename was stated.

Categories on Lexical Conventions in Independently-Considered Filenames

The second of the four category sets still analyzes each filename independently of the files surrounding it in its directory listing. As listed in table 5, there are five basic categories of different naming practices. These categories of analysis distinguish themselves as a group from the other nineteen because they require focusing on each substring as a lexical unit (or units) within the filename. Filenames in which multiple words were represented, using both truncation and full words for the substrings were also represented in the sample.

TABLE 7.
Second Category Set

Category	Proportion of Filenames Overall (%)	Proportion of Total L.E. ^a Filenames (%) ^b	Number of L.E. ^a Contributors	Proportion of Total Exp. ^a Filenames (%) ^b	Number of Exp. ^a Contributors
Full Words All ^c	41.8	45.0	3	40.0	6
<i>Abbreviations</i>					
Truncation	20.0	35.0	3	11.4	3
Acronym/Initials	9.0	10.0	1	14.3	3
Truncation, Compound, and Full Words ^c	9.1	2.0	2	2.9	1
Standard Contractions	5.5	0.0	0	8.6	3
Invented Contractions	3.6	0.0	0	5.7	1

^a Randomly-selected filenames created by the “Less Experienced” and “Experienced” participants.

^b Percentages are based on the total filenames contributed by all the participants in this experience-level. For the Less Experienced group, N=20; for the Experienced group, N=35.

^c Filenames containing index expressions, in addition to full words are included in these figures.

Each of these categories is differentiated from the others by virtue of the nature of its effects on the structural aspects of each of the substring units from which a word may usually be inferred. The semantic aspects of the substrings are considered in this part of the analysis only inasmuch as it is necessary to determine the words they represent. In other words, how a filename is classified depends on the how many (or which) characters are present in its inferred words but missing from each substring that represents each inferred word. For example, the substring “form” in the filename “form1terms.doc” would belong in the category called “Full Word,” which is the absence of the abbreviation effected by the procedures of the other four categories. The substring “cov” in “Resumecov2,” which the participant identified as his “resume cover letter,” results in a classification of “Truncation,” because it was formed by dropping one or more letters from the end of its inferred word, “cover.”

The next two categories belonging to this group feature abbreviation strategies which involve the creation of a “contracted” form of a word by deleting one or more letters from its interior rather than its very end. A filename belonging to the “Standard Contraction” category contains one or more substrings in a contracted form that follows the lexical conventions of abbreviations found in written English. The filename “wkcite.doc” exemplifies this with the abbreviation of “wk,” which is listed as a standard abbreviation for “work” in most English language dictionaries. In contrast, the “Invented Contraction” refers to the file creator’s omitting selected letters from a substring, resulting in a final product of two or more letters that is not a conventional abbreviation for the word it is

representing. “cometlgt.jpg” contains the substring “lgt” which appears to adequately suggest the word “light,” but does not match the conventional abbreviation, which is generally “lt.”

The fifth and last of this group’s categories, “Acronyms/Initials” can be viewed as the most extreme form of truncating, in that each substring representing an inferred word is missing all but its initial letter so that the substring is a single character long. Frequently the filename to which this has been done may contain a series of adjacent substrings, all truncated in this way, which forms an acronym. This can result in a gray area for classification, if the acronym, such as the “URL” in the filename “URLnotes.doc” has become so commonly used in the language, in place of pronouncing the words “universal resource locator,” that the combination of the letters ‘u,’ ‘r,’ and ‘l’ have nearly reached the status of a word in its own right, like “VCR” or “IBM.” In such instances, the filenames containing one-character substrings which combine to form a recognized term are simultaneously classified under the Acronym/Initials category and the Full Word or Proper Name (defined later) categories.

Full words.

As might be anticipated in a filenaming domain where long filenames are possible, especially among a group of PC users whose filename roots range in length from four to forty-six characters long, and average nearly twelve, the proportion of the randomly-selected filenames for both the experienced and less experienced groups that contain at least one full word was quite high at 82% (45 out of 55 files). The prevalence of this practice was slightly higher among the

less experienced participants' filenames (85%) than for those created by the more experienced (80%), but only very slightly; and all the individuals in both groups used whole words in at least two out of their five randomly-selected filenames. Nearly 42% of all the randomly-selected filenames with two or more substrings (48 in total) used only full words¹⁸ – and no abbreviations.

Three members of the less experienced group gave a single reason for their use of full words, that the application in which they were working offered it as a suggested filename. Said one participant about her file, "Abstracts & Indexes.doc": "It was in the field, which was fine with me. [...] I went with it." The fourth member of the less experienced group explained her use of full words as a habit, but qualified this with the condition that the full words not make the filename "too long." Three of the more experienced PC users offered visual clarity as a reason for using full words in some of their filenames, either for their own benefit : "[Nikeletter]...in two years, I want to know what was in the paper..." – or for the convenience of another person with whom they intended to share the file: "[canceragenda.html] I have to send everything to my boss, so I try to make everything incredibly descriptive..." Two members of the experienced group used full words to indicate the importance of the file's content: "[InterviewData.xls] "If it's an important document, I will spell it out...like 'InterviewData' was something very important; it wasn't a little snippet of data." Another individual used the Full Word approach in lieu of another naming scheme using initials that she had to abandon when it ceased to allow adequate differentiation between it and other related files: "[davidxxx2.rtf ¹⁹] I have a

number of relatives whose first names also begin with 'd' ... and so that's why I used his whole first name."

Truncation.

It is important to note that most of the participants did not tend to distinguish between a "contraction" and a "truncation," often using the two words interchangeably or, even more frequently, just using the term "abbreviation" to refer to both. Fortunately, using the evidence of the inferred words each substring was meant to represent and three different English dictionaries, determining into which categories each filename belonged became a relatively straight-forward exercise.

Again, due to the luxury of long filenames, the Truncation approach to filenaming is not as essential as it once was. However, it was still far from rare in this sample, characterizing around 20% of the randomly-selected filenames overall. The filenames created by the less experienced users exhibited a much higher proportion of truncation use (35%), with three of the four members contributing, than among more experienced participants' files (11%), with only three out of seven members contributing.

Four experienced users indicated their use of the truncation approach was fueled by a desire to limit filename length. One experienced individual had an almost aesthetic concern over the neatness of their file listings, "[LIBLIT.TXT] I didn't want to type out 'library literature' and have that whole filename stuck up there." Three others representing both experience levels were more concerned with reducing the number of keystrokes involved. Five out of the six individuals

who had at least one filename containing a truncated substring²⁰ were following a self-imposed length limit. For two of these, limiting their file lengths appeared to be a way of protecting themselves from the ravages of older systems not designed to handle long filenames:

[blusat.jpg] ... I started off in a Windows 3.1 environment, so I'm used to using eight characters to name my files. So even if I have the option to name it ... whatever I want to, almost always I will use a short name, *if I can come up with one that is meaningful to me* [italics added].

The caveat to this person's preference for short filenames, is consistent with other indications that adequate labeling of their files may still be the higher priority among many users of truncation.

Ten of the eleven files containing truncated words as substrings were multiple-word filenames, and seven of these also included at least one full word. Of the five randomly-selected filenames in which contractions were used, one was in excess of eight characters long, and the other four names were contributed by two individuals who were fairly consistent about limiting their filenames to an eight-character root or shorter. None of the less experienced participants used contractions, standard or otherwise, in any of the filenames observed in this study.

Standard contraction.

Of the three who did have a filename in which the contraction was a standard form of abbreviation used in written English, only one individual offered that fact as a justification for his choice to reduce "general" to 'genl,' calling it a "relatively common" abbreviation. All three mentioned "clarity" as a determining

factor in their decisions whether or not to contract a substring. One individual told how this influenced his decision to use contraction on only one substring out of the five in one of his filenames:

[FinalRptGWcover.doc] ...'r-p-t' was pretty standard for me. I wouldn't have a problem recognizing that. If I had tried to truncate 'final' or 'cover,' I think I would have had a harder time, but 'rpt' was pretty clearly 'report' to me.

Invented contraction.

The sole contributor of invented - or unconventional - contractions, was one of the experienced computer users who consistently limited his filenames to eight characters or less. In deference again to that self-imposed length limit, he shortened a substring using contraction. In two filenames his efforts resulted in invented contractions, which he indicated were generally acceptable even though "...it does make remembering things difficult." His creations, "lgt" and "drk" in the filenames "cometdrk.tif" and "cometlgt.jpg" illustrated his point that a phonetic similarity between the substring and the original word will aid recall. In other words, as he put it, "[p]ronunciation helps."

Acronyms or initials.

Four individuals, one of them from the less experienced group, made use of acronyms or initials in one or two of their five randomly-selected filenames. Overall, employing just the initial letter from one or more of the filename's inferred words, was used in seven out of the fifty-five interviewer-selected files, a usage rate of 9% over both groups. Generally, multiple inferred words that were turned into acronyms for the filename, usually are expressed as acronyms in many other contexts. The expressions "HSL" (Health Science Libraries), "URL"

(Universal Resource Locator), and “INLS” (INformation and Library Science), are just as easily interpretable to the students using them as the fully spelled-out terms they represent. For this reason the participants’ employment of these initials or acronyms in a filename could arguably be considered use of full words - or proper names.

Among those participants who used acronyms, the reason most often articulated was to keep the length of the filename shorter in keeping with a self-imposed file length limit. One individual, at the less-experienced level, did not express a concern over filename length but believed her use of an initial in one of her filenames was merely due to habit. She had named a file with her own name followed by her middle initial, exactly as she had been accustomed to signing her signature: her first name followed by her middle initial.

Categories on Semantic Properties in Independently-Considered Filenames

As the level of analysis becomes more abstract, a filename’s membership in a particular category is sometimes not easy to determine. The concrete nature of the properties observed at the more granular levels of filename anatomization makes classification more clear cut than what is often the case when one focuses on the meaning of the words represented in each filename. Due to the prevalence of multi-word filenames, it is possible for a single filename to simultaneously belong to all the categories in this set. (See Table 8 below.)

TABLE 8.
Third Category Set

Category	Proportion of Filenames Overall (%)	Proportion of Total L.E. ^a Filenames (%) ^b	Number of L.E. ^a Contributors	Proportion of Total Exp. ^a Filenames (%) ^b	Number of Exp. ^a Contributors
Function of Contents ^c	50.9	55.5	4	48.6	7
Function, no Subject ^c	34.0	33.3	3	34.3	6
Subject of Contents ^c	35.8	33.3	2	37.1	6
Subject, no Function ^c	18.9	11.1	2	22.9	5
Function & Subject of Contents ^c	17.0	22.2	1	14.3	3
Dates	0.0	0.0	0	0.0	0
Proper Name	36.4	35.0	3	37.1	7
Proper Name = Subject	14.5	20.0	2	11.4	4

^a Randomly-selected filenames created by the “Less Experienced” and “Experienced” participants.

^b Percentages are based on the total filenames contributed by all the participants in this experience-level. For the Less Experienced group, N=20; for the Experienced group, N=35.

^c Scoring for these categories were based on reduced totals, due to omission of two filenames from the counts because of insufficient data.

Subject of contents and function of contents.

Of the categories in the third set, determining the membership of the first two, “Subject of Contents” and “Function of Contents,” was the portion of this analysis most vulnerable to subjectivity. Generally, instances where participants chose to indicate the file content’s Function in relation to its companion files often involved the use of fairly generic words describing the file’s role, such as “assignment” or “paper,” rather than its topic of its contents. Those filenames belonging to the “Subject of Contents” category, contained inferred words (words represented by a substring in the filename) that specifically identified the area of knowledge, person, place, thing, or issue addressed within. The evidence upon which these classifications were based often required a certain degree of reading between the lines of statements made by individuals about a file, in an attempt to determine which of these roles the filename’s creator intended each inferred word to fulfill. In other words, within this study’s sample, the individual who

stated about his files, “Almost everything is named by function” during his interview was an exceptional case.

Perhaps asking more specific questions may have led to a higher prevalence of direct statements on filename meanings from the participants. However it was feared that such specificity could influence individuals to adopt underlying meanings to their creations which may not have truly been part of their original intentions. Consequently, most determinations of a file creator’s intent were based on indirect statements a participant made during the interview. A good example of this was one person’s comment regarding his file, “paper.doc.” He acknowledged that ‘paper’ was “not descriptive,” saying “I didn’t name it ... ‘year 2000 paper,’ which is the topic.” Such direct and indirect comments, in combination with the participant’s confirmation of the correct interpretation of the substrings in each filename, was used to justify classifying filenames as members of the “Function of Contents” or the “Subject of Contents” categories. Filenames lacking these two kinds of support, even if the meaning of the inferred words appeared obvious, were omitted from the scoring. Fortunately, of the fifty-five randomly-selected files encountered in this study, only two had to be excluded for lack of this kind of evidence.

Just over half of the fifty-three filenames scored in this portion of the analysis contained indications of their contents’ functions, while just over a third (35.8%) contained substrings representing the subject. The creators of nine files sought the best of both worlds by representing both the function and the subject of a file’s content in its name. More than one-half of the files named by less

experienced participants were functionally named, compared to slightly less than half of those named by the more experienced. The majority (66.7%) of the twenty-seven files evidencing use of the Function of Contents approach to naming contained no additional substrings to indicate subject; and of those filenames incorporating their contents' subject, less than half (47.4%) contained a reference to the file's Subject, without any indications of its function.²¹

For thirteen of the eighteen files named functionally, without any reference to Subjects in the name, most of the participants appeared to have no difficulties in recalling their contents because they were located in directories named for a subject, the course or some other context-providing name. For example, having placed his file "text.doc" into a directory he named "DissReport," one participant was able to confidently recall the contents of that file as the text for his dissertation based on its location in his directory hierarchy, even though the name he had given it was very generic. In two of these cases, the individual had only a short-term need for the file, with no definite plans to revisit it after a specific period of time:

[Memo3] I knew when I took the class what the different memoranda were and what they were on. And now ... I can't remember what Memo three is, as opposed to Memo two ... but it was very clear in class.

A single individual from the less experienced group also reported these motivations behind her rather generic names for some of her files, but combined them with the fact that each of these filenames had been application-suggested. That is, the application had drawn several words from the first line of the file's

contents and placed them in the filename field of a Windows “save as” dialog box.

Overall, the proportion of files whose names contained references to their contents’ subjects – either solely subject or a combination of substrings representing the both the file content’s function and subject – was lower than one might have anticipated. Only thirty-six percent of the classifiable fifty-three filenames incorporated any reference to their contents’ subjects. Less than a fifth of the files relied on subject alone for their names. It was very common for participants to volunteer the subject of a file’s contents during their interviews.

Oftentimes, the first comments an individual would make about a file chosen by the interviewer for discussion would be regarding its contents; and these statements would invariably include information about its subject. However, direct indications from participants of their reasons for incorporating the content’s subject into a file’s name were rare. Usually, any reasons explicitly mentioned were related to the broader context of the file’s intended use or audience. For one individual, it was important to include the subject *and* function in some of her filenames because she shared her files with her boss and found it useful to make these files’ names “incredibly descriptive.” Although this study scope did not include the characteristics and structure of the participants’ directory structures, many instances were noted where information about the subject of a file was contained in the directory name rather than the filename.

Just as one might suppose that the omission of subject in the filenames contained within such directories may have been to avoid redundancy, an

underlying factor motivating an individual's decision to include information about the content's subject in a filename could be his expectation that the filename would be accessed or stored outside its home directory. Viewed out of context, a filename with no representation of its content's topic would not be a good aid to recall or retrieval. Other possible reasons for much of the use of subject references in certain filenames could be to differentiate them from other files in a particular directory who's contents fill the same functions but for different projects.

Dates and proper names.

Determining which of the participants' filenames evidenced membership in the two remaining categories of naming approaches was much more straightforward a task than the analysis for the previous two categories and their subcategories. None of the randomly-selected filenames included any substrings or numerals that represented a calendar date. One individual brought up one instance of this on her own during her interview when asked about her file renaming practices. She pointed out a file called "july.wpd" which had formerly been part of a series of monthly reports. She had used the months of the year to improve their accessibility, as they were files she frequently referred to and shared with co-workers.

There were no examples of individuals using numerical dates in any of their filenames, which one might imagine might be preferred as a means of enforcing a chronological order on a set of such files in a file listing. Numerical dates used in a filename would be a compact means of adding information on file

versions to the name without exceeding a self-imposed length limit.²² A larger sample may well have included examples of filenames containing dates, so that these and other possibilities could be investigated further.

Among both the less experienced and experienced groups, use of a proper name in filenames occurred in more than a third of the interviewer-selected examples. Of these twenty filenames, the proper names represented in nine of these also doubled as their contents' subjects. For the rest, the proper name represented in the filename was that of an individual or institution related in some way to the file's contents, often as the intended recipient or source.

Roughly half of the filenames containing a proper name not also serving as the subject of the file's contents belonged to "throw-away" files, intended to fill a temporary function, rather than serve as a final product. The individuals expressed their use of the proper names in these situations as decisions without much thought beyond a desire to use a "unique" word that they would be able to locate shortly after saving the file, when it was time to either print out or otherwise put to use the information contained therein. An example of this use of a proper name was a file named "Adam1.doc" described by its creator as "not that important," as it was the text for an email he had written at home and saved to send the next day via his internet connection at his office. His only motivations for using "Adam," the intended recipient of the email, was to differentiate it from other files on his diskette at the time and aid him retrieving the correct text for his email the next day.

As for the other half of files incorporating proper names that were not their contents' subjects, their creators did not express any particular reasons for their choices. On at least one occasion, a less experienced individual who had three files she had named after herself, using her first name and middle initial followed by an index expression, gave her reason for this choice as another suggestion from the application she was using. The prevalence of this as a reason for her filename choices generally, suggests that perhaps this individual relied more on her recollection of file contents than on any set of naming strategies. Judging by the comparatively low volume of files she had in her directory at the time (only ten), this is not an unreasonable assumption.

Categories on Properties in Filenames Considered In Context

Unlike the previous three sets of categories, the six categories (as listed in Table 9 below) in the fourth category set naturally fall into three subsets which focus on the same phenomena as three rule-scheme types observed by Carroll, which he called "literal," "nonliteral" and "indexed" rule-schemes. The first three categories, labeled "Isomorphism," "File Type/Extension" and "Character or Substring Serial Position," are analogous to Carroll's "literal" types of rule-schemes, in that they are concerned with redundancies in the appearance of two or more filenames. "Isomorphism" refers to the practice of having two or more filenames with characters or substrings in common. Such an approach is exemplified in the companion files "usefullinks.doc" and "usefullinkscontent.doc." The Isomorphic approach to filenaming relates to the other two in a hierarchical fashion, in that they are variations of the same theme. Files named identically,

except for different three-letter extensions, are classified as evidence of the “File Type/Extension” approach to defining filename sets, whereas the “Character or Substring Serial Position” category refers to instances where a substring or characters common to multiple filenames varies in position within the name from one file to the next.

TABLE 9.
Fourth Category Set

Category	Proportion of Filenames Overall (%)	Proportion of Total L.E. ^a Filenames (%) ^b	Number of L.E. ^a Contributors	Proportion of Total Exp. ^a Filenames (%) ^b	Number of Exp. ^a Contributors
Isomorphism ^c	47.3	55.0	4	42.9	7
File Type/Extension	7.3	5.0	1	8.6	2
Character or Substring Serial Position	0.0	0.0	0	0.0	0
Index Expression	30.9	30.0	3	31.4	6
<i>Non-Isomorphism</i>					
Semantic Contraries	7.3	5.0	1	8.6	2
Semantic Relates	20.0	0.0	0	31.4	6

^a Randomly-selected filenames created by the “Less Experienced” and “Experienced” participants.

^b Percentages are based on the total filenames contributed by all the participants in this experience-level. For the Less Experienced group, N=20; for the Experienced group, N=35.

^c Scoring for this category was based on reduced totals, due to omission of seven filenames from the counts due to insufficient evidence for accurate classification.

Isomorphism.

A total of twenty-five of the randomly-selected files were identified as members of an isomorphic filename set. As before classification was determined by the combined evidence of a filename’s appearance together with its creator’s comments. Moreover, the majority of the filenames volunteered by the participants, as opposed to the randomly-selected files, were identified by each individual as members of such sets. Forty-seven percent of the fifty-five originally-chosen files were identified as members of an isomorphic set. A slightly higher proportion of the less experienced participants’ files (55%) were isomorphic set members than those created by the more experienced (42.9%).

Apart from the underlying reason of indicating that two or more files were related in some way, Isomorphism in this sample frequently appeared to be used as a way to force clustering on filename sets in an alphabetically-ordered file listing. The most basic reason for participants' use of isomorphism was to visually highlight related files. However, there were slight differences in the nature of these relationships within the twenty-five identified file sets which are important to note. Seven filenames whose appearance, along with their companions in the file listings, indicated apparent use of Isomorphism were excluded from this portion of the analysis. This was done because their creators' comments during their interviews did not sufficiently illuminate the nature of these files' relationships to other files.

Out of the twelve classifiable filename sets whose members were virtually identically-named (apart from the one or two characters of an index expression), the members of nine were different versions of the same content, and those of the other three sets were parts of a larger whole, such as a project or presentation. Of the ten classifiable filenames which were filename set members with only partially-matching names, the file contents of nine sets were components a larger entity.²³ In other words, the filename sets such as "ham.flo," ham2.flo," and "ham3.flo" usually meant that each contained a slightly modified copy of the others' contents. A file set more like "davidkxxx2.rtf," "dkhouse.jpg," "dkwindow," and "david.jpg," with only partial isomorphism, generally was a collection of pieces of some modular-type of entity, such as the following:

[market segmentation.html, marketsegmentation2.html, marketingaudit.html] Well, to look at [these] as an example - this is all marketing. ... they're all related and they build on each other. So 'market segmentation' is when we had to break up different markets into segments, and the 'marketingaudit' was when we had to audit the library. And they're all related ...

File type or extension.

This study's sample contained only a small handful of file sets in which the members were only distinguishable from each other because of their three-letter extensions. Overall, only 7.3% of the randomly-selected filenames were members of these kinds of file sets. In all of these cases, participants indicated that the contents of the files in each of these sets were virtually identical, apart from the differences that would be prescribed by virtue of their being different file types. An exception to this would be one pair of files, "usefullinks.doc" and "usefullinks.html," which involved further significant content changes made in the html file by the individual directly so that each one file was more a rough draft of the other than merely another copy with a different format.

Character or substring serial position.

There were no examples of filename sets wherein the names of the members matched in every respect save the serial order of their substrings. Reasons for this could be because such an arrangement may not as readily offer the visual and organizational advantages of clustering and may obscure the visual cues that a consistent substring order from one filename to the next affords. Additionally, differentiating two otherwise identically-named files with only the serial order of their respective substrings does not communicate

sequential or chronological order as effectively as other naming approaches, such as the incorporation of index expressions or dates. In Carroll's study, he does note a single case where the serial position was used to differentiate between related files, although it involved just the use of a repeated index expression used as a prefix in one file "XTIMER" and a suffix in its companion "TIMERX" (1982, p. 416).

Index expression.

Use of an Index Expression in a filename involves the addition of a character or number that carries a distinct meaning or purpose. The use of an Index Expressions was a filenaming practice that frequently co-occurred with the participants' use of Isomorphism in filename sets. This is easy to understand in light of the fact that no two files sharing the same directory can share the same name and extension. Adding an index expression appeared to be a popular way to distinguish between files with related content. Every single use of an index expression among the originally-selected fifty-five filenames occurred in files which were members of an isomorphic file set. The one exception to this ("Adam1.doc") stood alone only because, according to the creator, its companion files were all deleted for disk space prior to the interview. Overall, use of an indexing scheme characterized under one-third of the identified filename set members in the sample.

The specific organizational purposes fulfilled by the index expressions used by participants in their filenames were quite varied. Of the seventeen filenames employing index expressions, only four lacked the accompanying

evidence of their creator's words to confirm that their primary purpose was to differentiate the filename from that of a companion member of the set. The second most prevalent reason for employing an indexing strategy in filenaming was a desire to indicate file versions. This was used almost exclusively by the less experienced computer users. At the same time numerical or letter index expressions often were relied on for chronological information, usually to indicate which of the versions was most recent: "[DonnaA.doc, DonnaA2.doc] ...one of them is the first draft. This [indicating DonnaA2] is the second draft."

Of the five occasions of index expressions being used to indicate parts of a whole, all were filename sets created by experienced computer users. One of these featured two indexing schemes, numbers and letters, with the letter scheme embedded within the number scheme: "DIALOG2.doc," "DIALOG2A.doc," "DIALOG3.doc." On two occasions numbers were embedded in a filename for no apparent reason, but turned out to be the abandoned beginnings of another indexed filename set: "[form1terms.doc] ...the reason it says 'form one' is because I... [thought] I would do separate ones, but I ended up consolidating it."

Although it appeared to be the apparent motivation behind many more of the occasions of indexing, only one individual explicitly mentioned his use of index schemes to enforce the order of the filenames in the file listing. One of these was characterized by the use of numbers and decimals to construct an two-tiered system, which he borrowed from the traditional arrangement for most outlines found in the domain of written English: (e.g. 3. Sources.ppt, 3.1

Sources.ppt, 3.2 Sources.ppt, etc.]. In another filename set, he pointed out his practice of prefixing his single-digit numbers with a zero, so that when an indexing scheme required the use of numerals above 9, the zeros would enforce numerical order on the file listing, (e. g. 01, 02, 03 ... 10).

Semantic contraries and semantic relates.

The last two categories of the fourth set of filenaming strategies ignore the morphological features of filenames and instead seek to discover patterns solely within the semantic properties of the substrings found in a set of filenames. The concept of structural redundancy is abandoned in favor of links based on the meanings of words - or, as Carroll calls it, "conceptual redundancy" (1982, p. 415). Evidence of the use of "Semantic Relates" is when one or more substrings in a filename represent words which have a conceptual relationship with the inferred words from another filename. A good example of this was the pair of files named "carpenter.doc" and "muir.doc." The participant indicated that both of these files contained articles with very long titles, and so she decided to name them with the authors' names, hence relating them conceptually. A slight variation on this theme was exemplified by other participants' use of conceptually-linked terms to designate the components of a paper (e.g. "wkcite.doc" for "works cited" and "paper.doc" for the "paper" itself) or the portions of web site (e.g. "mod1index.html," "mod3line.html," and "mod2page.html"). Use of these tactics was limited to filename sets created by six of the participants from the experienced group, but were present in nearly a third of their thirty-five interviewer-selected filenames.

The logical opposite to Semantic Relates, labeled in this analysis as “Semantic Contraries,” features the practice of conceptually relating two filenames by using inferred words with opposite meanings. The use of semantic contraries among the interviewer-selected filenames, which were members of filename sets, was not very high. Only four filenames among the fifty-five selected by the interviewer were members in a filename set linked by semantic contraries, and two of these were part of the same set. All of the filenames were multiple-word names; one set was contributed by a member of the less experienced group; and the other two were created by more experienced users.

Unfortunately, none of the participants’ comments spoke directly to the question of why this particular approach was employed for these filenames. A reason that was implied, in the case of the file set, “cometdrk.tif” and “cometlgt.tif,” was that the inferred words “light” and “dark” were used to distinguish between two images of a comet which had little to differentiate them in appearance besides the fact that one was a little brighter than the other. It could well be that the element of relationship is not truly contributed to by the semantic contraries, but is solely the function of the isomorphistic use of “comet.” Had the noticeable attribute of *one* of these been its size instead of its relative darkness, the names could well have been “comet light” and “comet small;” and “small” and “light” are not semantic contraries. The same difficulties apply to the other two sets, “Astr-night1.doc” and Astr-day1.doc;” and “Assignment 2 - Final Submission” and “Assignment 2 - Initial Submission”²⁴

Just from observing the data that was collected on these filename sets, one could speculate that, in addition to a chief reason of showing the relatedness of two files and distinguishing between them, the other functions of Semantic contraries in these cases could have been to indicate the parts of a whole entity, such as a project or an image collection. In the set where the contraries “final” and “initial” were used, one could argue that these inferred words were also sources of sequential information or, depending on the nature of the assignment, indicators of content versions.

Parts of Speech

The part-of-speech analysis of the data was based almost entirely on the classification structure outlined in the Carroll study (1982, p. 422- 423).

Filenames in this sample broke down into groups of Simple Nouns, Compound Nouns, Noun-Adjective Combinations, Noun Phrases, and Others. Probably because of the smaller range of file types, the verbs and verb phrases encountered in other studies were not present in this sample. The grammatical makeup of each filename was based on its creator’s identification of the correct inferred words. Only one filename lacked evidence to support a clear enough definition and had to be omitted from the analysis.

As in prior studies on naming, most of the filenames in this sample were nominal (68.5%). Of these the majority of these were simple nouns (e.g. “paper.doc,” “lynn.doc,” etc.), closely followed by compound nouns (e.g. “pathfinderpage.html,” “Anderson Consulting,” etc.). Only three of the filenames were noun phrases, exemplified by the filename, “Interview with Sergeant Lxxx

Bxxx.rtf.”²⁵ The proportions of the respective experience level groups’ filenames that were compound nouns were fairly similar.

The incidence of simple noun filenames among files created by the more experienced participants was slightly higher at 40% than what occurred among the collection of filenames made by the less experienced (26%). More than three times as many of the less experienced member’s filenames were noun phrases than what occurred among the thirty-five filenames from the experienced group members (2.8%). Only six of the fifty-four filenames analyzed for parts of speech were simple nouns modified by adjectives, and all of these were created by the experienced PC users.

The remainder of the files were categorized as “other.” Within this grouping were small clutches of filenames with unique characteristics worth mentioning here. The first of these, and the most numerous, were filenames which read as two or more separate phrases or labels. In some cases, there was evidence that the creators of these files were creating these longer names at some point to substitute for subdirectories. The comments of one experienced participant about one of these files (“FinalRptGWcover.doc”) provided one explanation of why this approach was taken:

I didn’t rely ... as much as I could have on the Glaxo directory name. I repeated that information in the name of the file. It’s “Final Report Glaxo Welcome,” and it’s the cover letter that goes with the final report. ...I probably could have named it just Final Report Cover-dot-d-o-c and left off the ‘GW.’ I’m not sure if I had it in another directory originally and moved it to this one... I think I did, actually.

A similar motivation was expressed by another participant, a member of the less experienced group, who contributed three of these types of filenames. However, in these cases, the multi-phrase filenames were her strategy for dealing with her inability to create her own subdirectories. In short, when the contextual information often provided by the directory name in which a file is located was for whatever reason absent, some PC users instead chose to represent this information in the filename, often as a visually separate unit of information.

Two other filenames shunted to the “Other” group might have counted as simple nouns modified by an adjective were it not for their syntax. “cometdrk.tif” and “cometlgt.tif” were defined by their maker as “comet dark” and “comet light,” rather than “light comet” and “dark comet,” which set them apart from the other simple noun-adjective combinations with a more conventional syntax, such as “genlinfo.doc” (“general information”), which was also created by the same individual. If there was a reason for this different syntax, it was not directly stated. The fact that putting the word “comet” first allowed the members of this file set to cluster in an alphabetically-arranged file listing was a likely explanation, but had to be considered speculation in the absence of direct evidence.

The last subgroup of the “Other” collection was occupied by a single file, “hell.doc,” Based on just the filename, the first inclination was to call it a simple noun. However, the sense in which the word was used by the filename’s creator, “Hell!” made it an interjection. According to its owner, this was a low-priority file, created for a temporary purpose, which is why she felt free to indulge her whim on this occasion.

Issues Surrounding Filenaming

When to Name

In addition to the discussions about individual filenames and filename sets, participants were asked about their general habits while working with their files. The first of these were regarding the point during working with a file at which each individual would name it. Because naming a file is a prerequisite to saving one's work on the hard drive and getting it out of the vulnerable location of temporary memory, most persons' decisions to name a file at particular time were driven by their need to save it. As shown in Table 10 below, participants' answers tended to fall in one or more of the same three categories: "Almost Immediately," "After Some Work or Typing," and "At or Near the End."

TABLE 10.
When Participants Name Their Files

Experience Level	"Almost Immediately"	"After some work/typing"	At/Near the End
Less Experienced	0	4	0
Experienced	2	4	1
Both Levels	2	8	1

Only two of the experienced PC users said their usual habit was to name their file "almost immediately," with one person claiming to name and save a file "before I even type anything." However, the majority of individuals (All four less experienced and four experienced) preferred to wait until after they had been working a short time. Several characterized this short time as the first fifteen to twenty minutes of typing or the point when they were ready to take their first break. Still others indicated they demarcated this short time with their first interruption or with the point when what they had just typed began to seem

“significant” or difficult to re-create should it be lost. One person qualified his answer with a few criteria, indicating that generally he did wait to name his files after a relatively short period of time working on them, except when they were “short” and “simple,” in which case he would often wait until the end. He also mentioned some Windows applications, such as Access, which require naming one’s file before it is possible to add any content. Only one individual claimed that he generally waited until he was nearly done creating the contents before deciding on a name for the file, and only saving earlier than this when his work was interrupted.

Renaming

In addition to the timing of their filename decisions, participants were asked whether they were in the habit of renaming their files, and if so to describe the reasons for which they would do so. Due to some missing portions of the interview transcript, one participant’s response to this question was not available for analysis. Of the remaining ten, two of the less experienced group and one experienced participant indicated that they were not in the habit of renaming their filenames. Among those whose custom was to rename their files on occasion, the most frequently-given circumstance was when they noticed a filename was not adequate. This would be either because it was not meaningful enough because it had been named in haste or was an application-suggested name, or had not been a sufficient aid to finding the file.

Two individuals from the experienced group mentioned specific occasions where they had re-opened a file to change or add to its contents, and then

renamed it to reflect this. A similar scenario was described by two others, where the deletion or addition of some files in a directory compelled them to rename several members of a filename set either to maintain differentiation and adequate descriptiveness or to eliminate no-longer-needed substrings. Another three participants from the experienced group talked about occasions where a technical problem encountered when transferring a file from one system to another forced them to make changes in its name.

Retrieval

As described earlier, at the end of each interview there was the short exercise presented as a test of each participant's ability to retrieve the source of a printout of one of his files' contents. Each student was allowed three attempts, although none were allowed to open any of the files to determine their success until after all three guesses were made. This short test was only conducted a single time with each participant, except on those occasions when a second one was needed because the initially-selected file was found to be named by someone else. (See Table 11 below.)

The purpose of this exercise was more to gain insights into the each participant's process of interpreting his filenames for information retrieval than as a measure of the effects of his file system on retrieval success. Each person was asked to point to three files which he felt might contain the source of the printed contents, with the first being the most likely choice followed by his second and third attempts.

TABLE 11.
File Retrieval Exercise- Success Rates

Experience Level	First Attempt	Second Attempt	Third Attempt	Failed to Find Fileame
Less Experienced	3	-	-	1
Experienced	5	-	-	2 ^a
Overall	8	-	-	3

^a Two individuals found the correct names, but not the correct files. One individual had two files identically-named, except for the extension. He chose the one with an extension different from the that of the file from which the printed content was obtained. Another individual chose a file whose content was identical to the printout's, but it was not the correct file, only a related one inadvertently placed in another subdirectory.

Three individuals failed²⁶ to identify the source of their printouts. All those who correctly identified the source of their printout did so with their first attempts, except for one individual who was correct with his second choice. For the most part, the second and third opportunities for alternative choices were never actually needed by almost all who retrieved the correct file. In describing how they knew from which file the printouts originated, three of the successful eight directly referred to the subdirectory name as a means of “homing in” on the correct file. “[Final.doc] It’s one of these four documents because it’s in the information Model’s directory” One of the individuals who did not find the correct file was able to at least identify a branch in his hierarchy of subdirectories where it was located. “[Fammhome.doc] The part that tells me it’s a homepage [‘home’] tells me it’s in ‘webstuff’ [indicating one of his subdirectories].” Once he identified the correct subdirectory, another participant relied on the file type to find the source of his printout:

[URLnotes.doc] I knew it was going to be in the file [sic] for Special Warfare Center & School (SWCS) and... it's the only doc file in there. All the rest are PowerPoint. I just looked for the little 'W' symbol.

Three more individuals were able to rely a great deal on their recollection of having named the contents in the printout, indicating that they had created the file less than a few weeks before the interviews. One member of the less experienced group who was able to find her file was in the habit of creating unusually long filenames (average length: 30 characters), and so was able to easily identify the correct file simply by matching each of the filename substrings with words found at the top of the first page of the printout.

Summary of Observations

Even from this relatively small sample, a very wide variety of filenames, exhibiting many different strategies adopted by their creators, were encountered. There was at least one example among the fifty-five randomly-selected filenames for twenty of the twenty-two distinct filenaming practices identified in the analysis portion of the study. One of the two categories lacking specimens, "Date," which refers to the practice of incorporating a calendar date into the filename, was based on a participant's account of his using this approach for a file that was not accessible during the interview.

Additionally the categories within each group were far from exclusive. Several co-occurrences of two or more naming strategies in the same filename were observed. Indeed some strategies appeared to be employed in conjunction with certain others more frequently than they were used on their own. In a larger sample, further quantitative analysis may support these speculations on the

existence of additional categories which feature two or more distinct filenaming tactics operating in unison, what Carroll has called a “meta-rule-scheme” (1982, p. 417).

Overall, individual filenames averaged simultaneous membership in five of the twenty-two different categories of approaches to filenaming. There was some difference between the files of the less experienced participants and those created by more experienced PC users in this respect. Among the less experienced group, the most common number of membership in filenaming practice categories for a single file was four, with the median at four-point-five. Among the filenames created by the more experienced, the most common number of simultaneous classifications for an individual filename was seven, with the median value falling at six. This seems to harmonize well with the common sense expectation that an individual with more PC experience employs more sophisticated and involved filenaming strategies than a novice user, if for no other reason than he generally must work with a larger accumulation of files in his directories.

Profiles of the Typical Filename

The profile of the most common filename encountered during the study was the type produced by the experienced PC users, which is understandable since experienced participants outnumbered the less experienced by three in the sample. The typical filename had a length of around nine or ten characters, and its substrings usually were all lowercase characters, although some capitals were occasionally included in order to separate the multiple substrings and increase

visual clarity for easier reading. Spaces and symbols were generally not used in this more prevalent type of filename, although they were more frequently encountered in filenames where capital letters were used. The individuals creating such filenames were generally cognizant of the implications of straying too far from their self-imposed length limits and their avoidance of spaces and certain symbols. Despite this concern with filename length and characters that will work with older software and systems, there was also a desire for filenames which were easily interpreted. Consequently, the typical filename contained at least one full word, along with one or two acronyms which did not impact the filename's length as much. One suspects that, apart from these external sources of concern, the experienced user often employed filename shortening strategies to not only save on typing, but because it was a habit to which he had become accustomed from years of working with the eight-character restriction.

The inferred words represented in the typical filename indicated the function of its contents, while only occasionally being subsidized with a substring representing its subject. If the subject was indicated at all, the experienced PC user generally used his directory names to provide the context of a subject for his functionally-named files. Typically, reasons for those rare instances when the subject was represented in the filename were usually that it was a situation where the file's presence in an appropriately-named subdirectory was absent, such as when the file was to be share with others or when the directory simply had not been created.

When there were other files related to the typical filename, this relationship was generally shown via matching substrings, always in the same serial positions within their respective filenames. Differentiation between the members of a typical filename set in which the members were visually linked via morphological redundancies were either achieved with the addition of another substring or index expression - or both. In these cases, differentiation was required, not to tell one version of a file from another, but to distinguish the separate parts, like modular pieces of a whole product. In fact, it was atypical for the experienced user in this sample to keep earlier versions of a file at all, once the final product had been completed.

The typical filename created by the less experienced users was twice as likely to use a mixture of capital letters and lowercase, than lowercase alone. However, this generally was more in deference to the conventions of written English than as a means to demarcate the substrings within the filename. Demarcation of the substrings was frequently achieved through use of spaces or even punctuation. Frequently, these characteristics originated in the filename suggested by a Windows application, which the less experienced user often adopted with few if any modifications, although the desire for enhanced legibility and plain old habit were other motivations. When abbreviations were used, they were almost exclusively truncations instead of standard or customized contractions. The reasons for their use reflected the less experienced user's desire to reduce the number of necessary keystrokes more than any concern over file length.

The Less experienced user's typical file contained representations of the content's function rather than subject, but usually because the individual had only a short-term interest in the file and so was less inclined to go to the trouble of working out an appropriate subject term with which to label it. Alternatively, possibly further reflecting the preceding short-term interest, the typical filename created by a novice was at times just someone's name, either the creator's or the intended recipient of its contents. When the file in question was also related to other files in the same directory (almost always as different versions of the same content), a partial match between the members of this set was sought using isomorphism combined with a single index expression (generally a number), used to differentiate between them.

Concluding Remarks

Future Research

The contrasting of practices and motivations between experienced and less experienced individuals observed here suggests several differences which should be investigated using a larger, randomly-selected sample. It is important to remember that the observations gleaned from these eleven interviews are only the experiences of a small group of volunteers, and so cannot be used to characterize any sort of universal type. However, the insights gained from the in-depth examination of this small portion of the universe should be a useful source of information for building tools of future study; these studies could anticipate the range of tactics and motivations in the individual's management of his personal computer files. Further research using quantitative measures will make possible

more sophisticated analyses of the interactions between the categories of filenaming approaches identified in this study. It will also provide a more current understanding of filenaming behavior among the members of other populations. Further research of this kind promises a greater understanding of the processes employed by many different groups of computer users. Perhaps designers of computer file management systems could use these insights to revolutionize the process of labeling one's personal files for future retrieval. However, it is important to keep in mind the changing contexts within which filenames are created that we have witnessed since the advent of networked computing and the Internet. These trends suggest that file management systems geared to just the individual are an evolutionary dead end.

The Internet, and workgroup software, such as Lotus Notes, which facilitate collaborative work and file sharing, may have made filenames intended for the consumption of others much more prevalent. A filename or directory name that is only meaningful to its creator may not be good enough when it is part of a URL (Universal Resource Locator) or a project worked on by others.

Future Information Science research should reflect the increasing prevalence of these trends in computer file use. The next step for further study in this area could focus on files named for wider audiences. It would be interesting to assess the extent to which the patterns and motivations observed here exist among groups working together with shared files, or in an analysis of filenames found on the Internet.

Implications for Basic Training in Computers

In light of this broadened exposure to filenames, perhaps what we discover about the filenaming habits of individuals would be more profitably applied to computer training. An overview of approaches to filenaming and organization, together with their various tradeoffs and implications, should occupy a more prominent place in basic computer training curricula and texts. During the course of this study, the data revealed several instances where individuals employed naming practices which eventually led to various technical or organizational difficulties that might have been avoided with more extensive file management training. Left to their own devices, many computer users (particularly beginners) find themselves playing the role of a magician's apprentice, churning out computer files like so many enchanted broomsticks, with hardly any guidance on how to label and manage them, let alone share them with others. Such individuals would likely benefit from a guided tour of the filenaming experiences, and mistakes, of other computer users.

NOTES

- ¹ Users of other operating systems, such as IBM's OS/2 (released in 1993), the Mac O/S and all Unix systems, had access to long filenames all along. But these operating system are not widely used by average PC owners.
- ² "8.3 version" refers to filenames with a root of eight or less characters followed by three-letter extension. The root and the extension are separated by a "dot" or period.
- ³ At times it may be truncated to fewer than the first six characters if this is necessary to achieve uniqueness.
- ⁴ Or from an MS-DOS window inside Windows 95.
- ⁵ This is not meant to imply that it is impossible to accommodate non-standard extensions, as the latest versions of Windows do allow the user to manually assign parent applications to a customized filename extension.
- ⁶ Two hundred and fifty-five is actually the upper limit for Windows 95, which uses VFAT to enable longer filenames (See Prorise, 1996). Unix, OS/2 and Windows NT use different file systems. They all allow even longer filenames and have different restrictions on what characters may be used.
- ⁷ When the random number selection process resulted in a subdirectory being chosen, that subdirectory would be opened and another random number was generated until a file was selected from within it.
- ⁸ The interviewer printed directly from the file listing; and although the contents of these files were very briefly flashed up on the screen, due to the need for Windows 98 to open the file in its parent application in order to format the pages for printing, the interviewer did not look at the contents of any file. This was in accordance with the privacy measures agreed to by each participant.
- ⁹ If the filename appeared on the printout or the file was not one created by the participant, another file was selected instead.
- ¹⁰ The five filenames include the four initially selected during the first part of the interviews, plus the file whose printed contents were presented to the participant for the retrieval exercise.
- ¹¹ The follow-up interviews employed by Carroll dealt with correctly interpreting the inferred words which the eight-character filenames represented, and not to discover the specific reasons the creator adopted a particular naming approach. Although this type of information may have been volunteered by the participants during these interactions, it is not presented as support for any speculations on motivations made by Carroll.
- ¹² This would be impossible to avoid completely because of the interviewer's own filenaming experience and the anecdotal evidence absorbed in daily contact with other computer users, not to mention the influence of insights gained during interviews conducted earlier.
- ¹³ An "index expression," is defined as the addition of a character or number to a filename that carries a distinct meaning or purpose.
- ¹⁴ Carroll's miscellaneous category included nonstandard file types created by the individual.
- ¹⁵ Carroll introduced the term "inferred word" to indicate the word that a filename creator intends a substring within the name to represent. For example, "pap" in the filename "firstpap.doc" stands for the word "paper."
- ¹⁶ This does not include the three letter extension after the "dot" at the end of the filename's root.
- ¹⁷ Since the choice to keep or discard the application-created filename was made by these individuals, and some modifications (such as deleting portions of them) were made, these names were counted as filenames created by them.
- ¹⁸ This figure also includes filenames where index expressions (defined later) were present.
- ¹⁹ x's are substituted for some of the characters in the filename in order to ensure this participant's privacy.

-
- ²⁰ Actually, three individuals from the experienced group had at least one filename out of the five randomly-chosen during their interviews in which truncation was used. An additional three members of this group had at least two truncation examples among the filenames they brought to the attention of the interviewer themselves.
- ²¹ It could be argued that the three-character file extensions can indicate function; but these were not counted in this analysis due to the fact that participants rarely if ever have a choice over what their file extensions will be when working with many of the software applications used here.
- ²² This only works if single digits are prefaced by zeros (e.g. 01, 02, ... 10, 11, 12).
- ²³ Due to insufficient evidence, one filename was excluded from this analysis.
- ²⁴ As these two were Macintosh files, no file extensions were used.
- ²⁵ Original letters were replaced with x's as a privacy measure.
- ²⁶ It may be overly strict to refer to two of these three occasions where the participant did not select the correct file as failures, due to the fact that these two individuals did select files with contents that were virtually similar to those depicted on their printouts. On one of these occasions the name of the wrong file was identical to the correct file, except for a different extension.

References

- Beckman, D., & Hirsch, D. (1996, November). File it under EZ2REMEMBER: Adherence to naming conventions can avoid document retrieval problems. ABA Journal, 82, 87.
- Black, K. J., & Motz, A. A. (1989, January). Standardizing PC file names using DOS directories. ARMA Records Management Quarterly, 23(1), 14-17.
- Bott, E. (1995, October 20). Managing files. *ZDNet*. Retrieved July 30, 1998 from the World Wide Web:
<http://www.zdnet.com/pccomp/features/951020/feature2/sub3.html>
- Boyce, J. (1998). Inside Windows 98. Indianapolis, IN : New Riders.
- Carroll, J. M. (1981). Creating names for things. Journal of Psycholinguistic Research, 10(4), 441-455.
- Carroll, J. M. (1982). Creative names for personal files in an interactive computing environment. International Journal of Man-Machine Studies, 16, 405-438.
- Carroll, J. M. (1985). What's in a name? : An essay in the psychology of reference. New York: W. H. Freeman.
- Carter, A., & Robinson, G. (1999, January 13). Windows 95/98: File Management. **Academic Technology & Networks**, University of North Carolina. Retrieved January 26, 1999 from the World Wide Web:
<http://help.unc.edu/cgi-bin/getdocs?docnumber=dws10&doctype=Desktop&platform=Windows&product=File%20Management>
- Collantes, L. Y. (1995). Degree of Agreement in Naming Objects and Concepts for Information Retrieval. Journal of the American Society for Information Science, 46(2), 116-132.
- Cramer, M. (1990). Structure and mnemonics in computer and command languages. International Journal of Man-Machine Studies, 32, 707-722.

- Davis, F. & Crosby, K. (1998). The Windows 98 Bible. Berkeley, CA : Peachpit Press.
- Dunn, S. (1998, April). Get along, little file name. PC World, 16(4), 272-274.
- Ellingson, D. (1995, June 26). Don't tear your hair when you can use your careware. Las Vegas Business Press, 19(25), 21.
- Goldsborough, R. (1995, November). Let Windows 3.1 programs use long file names. PC World, 13(11), 84.
- Halverson, M. & Young, M. (1997). Running Microsoft Office 97 : Standard and Professional (Vols. 1 - 2). Redmond, WA: Microsoft Press.
- Hardy, D. R. & Schwartz, M. F. (1996, May). Customized Information Extraction as a Basis for Resource Discovery. ACM Transactions on Computer Systems, 14(2), 171-199.
- Hewitt, S. M. (1994). Computers don't byte. Rocklin, CA : Prima Publishing.
- Hodge, M. H., & Pennington, F. M. (1973). Some studies of word abbreviation behavior. Journal of Experimental Psychology, 98(2), 350-361.
- Inter Trade Corporation. (1996). PC Guide: Introduction to Computers (4th ed.). Norcross, GA : ITC Publishing Group.
- Jamsa, K. (1995). Welcome to ... Personal Computers (3rd ed.). New York, NY : MIS:Press.
- Joyce, J. & Moon, M. (1998). Microsoft Windows 98 at a Glance. Redmond, WA: Microsoft Press.
- Katzenberg, B. & Piela, P. (1993, June). Work language analysis and the naming problem. Communications of the ACM, 36(6), 86-92.
- Lansdale, M. (1990). The role of memory in personal information management. In P. Falzon (Ed.), Cognitive Ergonomics: Understanding, learning and designing human-computer interaction (pp. 39-49).

- Lee, A. Y., Foltz, P. W., & Polson, P. G. (1994). Memory for task-action mappings: mnemonics, regularity and consistency. International Journal of Human-Computer Studies, 40, 771-794.
- Livingston, B. & Straub, D. (1995). Windows 95 Secrets. 3rd ed. Foster City, CA: IDG Books Worldwide, Inc.
- Magid, L. J. (1996). the little PC book (2nd ed.). Berkeley, CA : Peachpit Press.
- Malone, T. W. (1983, January). How do people organize their desks? : Implications for the design of office information systems. ACM Transactions on Office Information Systems, 1(1), 99-112.
- Mamis, R. A. (1994, August). File names that tell all. Inc., 16(8), 108.
- Minter, A. L. (1990, April). File naming on PCs. Management Services, pp. 20-22.
- Nielson, J. (1996, February). The Death of File Systems. useit.com: Jakob Nielson's Website. Retrieved January 26, 1999 from the World Wide Web: <http://www.useit.com/papers/filedeath.html>
- O'Hara, S., Vega, D. B., & Kelly, J. (1997) . Discover Office 97. Foster City, CA: IDG Books Worldwide.
- Person, R. & Rose, K. (1995). Using Word for Windows 95 (Special ed.). Indianapolis, IN : Que.
- Person, R. & Rose, K. (1997). Using Microsoft Word 97 (Special ed.). Indianapolis, IN : Que.
- Press, L. (1990, November). Personal Computing: Windows, DOS and the Mac. Communications of the ACM. 33(11), 19-26.
- Press, L. (1994, March) Emerging Dynabase Tools. Communications of the the ACM, 37(3), 11-16.
- Prosise, J. (1996, June 25). How Windows 95 stores long filenames. PC Magazine, 15(12), 217-219.

- Rathbone, A. (1992). Windows for dummies. San Mateo, CA : IDG Books Worldwide.
- Rathbone, A. (1993). OS/2 for dummies. San Mateo, CA : IDG Books Worldwide.
- Rathbone, A. (1995). Windows 95 for dummies. Foster City, CA : IDG Books Worldwide.
- Rathbone, A. (1998). Windows 98 for dummies. Foster City, CA : IDG Books Worldwide.
- Roberts, T. (1992, January). Learn to play the name game: Follow just a few simple rules, and you'll be able to find any file on your hard disk. Compute!, 14(1), 66.
- Robinson, P. (1995, December). The explorer files: discover how File Manager's successor can ensure your data never again becomes uncharted territory. Home PC, 2(12), 207-214.
- Taylor, S. J. and Bogdan, R. (1998) Introduction to qualitative research methods : a guidebook and resource. New York : Wiley.
- Taylor, S. J. and Bogdan, R. (1984) Introduction to qualitative research methods : the search for meanings. New York : Wiley.
- Tinker, M. B. (1995, Jan/Feb). When 8 isn't enough. Legal Assistant Today, 12(3), 78-80.
- Teasley, B. E. (1994). The effects of naming style and expertise on program comprehension. International Journal of Human-Computer Studies, 40, 757-770.
- Weverka, P. (1996). Word 97 for Windows for Dummies : Quick Reference. Foster City, CA : IDG Books Worldwide.
- Weidenback, S. (1986). Beacons in computer program comprehension. International Journal of Man-Machine Studies, 25, 697-704.
- Zoepritz, M. (1986). A framework for investigating language-mediated interaction with machines. International Journal of Man-Machine Studies, 25(3), 295-315.

Sample Screenshot of File List

Name	Size	Type	Modified
01_26_lit_review.doc	28KB	Microsoft Word Document	1/26/99 5:41 PM
01_26_references.doc	63KB	Microsoft Word Document	1/26/99 6:02 PM
01_28_lit_review.doc	42KB	Microsoft Word Document	1/26/99 6:04 PM
01_28_references.doc	44KB	Microsoft Word Document	1/26/99 6:03 PM
01_31_lit_review.doc	31KB	Microsoft Word Document	1/31/99 9:25 PM
01_31_lit_review.doc.bak.doc	31KB	Microsoft Word Document	2/4/99 4:58 PM
01_31_references.doc	48KB	Microsoft Word Document	1/31/99 8:54 PM
02_04_population_training.doc	36KB	Microsoft Word Document	2/4/99 4:59 PM
02_07_lit_review.doc	60KB	Microsoft Word Document	2/8/99 5:26 PM
02_07_population_training.doc	33KB	Microsoft Word Document	2/8/99 9:16 AM
02_07_references.doc	30KB	Microsoft Word Document	2/8/99 11:19 AM
02_08_lit_review.doc	36KB	Microsoft Word Document	2/9/99 8:55 PM
02_08_population_training.doc	52KB	Microsoft Word Document	2/9/99 3:09 PM
02_08_question.doc	30KB	Microsoft Word Document	2/8/99 12:37 PM
02_08_references.doc	32KB	Microsoft Word Document	2/11/99 6:57 PM
02_09_population_training.doc	38KB	Microsoft Word Document	2/9/99 6:35 PM
02_11_lit_review.doc	38KB	Microsoft Word Document	2/26/99 11:31 PM
02_11_population_training.doc	42KB	Microsoft Word Document	2/26/99 11:31 PM
02_11_references.doc	32KB	Microsoft Word Document	2/11/99 7:10 PM
03_01results.doc	69KB	Microsoft Word Document	3/2/99 10:35 PM
03_03Figures.doc	44KB	Microsoft Word Document	3/4/99 8:09 PM
03_03results.doc	61KB	Microsoft Word Document	3/4/99 5:15 PM
03_04results.doc	88KB	Microsoft Word Document	3/5/99 7:06 AM
03_05Figures.doc	64KB	Microsoft Word Document	3/6/99 5:59 PM
03_05results.doc	103KB	Microsoft Word Document	3/6/99 5:50 PM
03_06discussion.doc	92KB	Microsoft Word Document	3/7/99 4:41 PM
03_06results.doc	130KB	Microsoft Word Document	3/6/99 9:14 PM
03_07discussion.doc	39KB	Microsoft Word Document	3/7/99 7:36 PM
03_07Figures.doc	36KB	Microsoft Word Document	3/7/99 2:53 PM
03_07population_training.doc	40KB	Microsoft Word Document	3/7/99 5:24 PM
03_07results.doc	66KB	Microsoft Word Document	3/7/99 6:07 PM
03_08_references.doc	48KB	Microsoft Word Document	3/7/99 7:19 PM
03_08Figures.doc	74KB	Microsoft Word Document	3/7/99 3:52 PM
03_08Figures_by_Group.doc	66KB	Microsoft Word Document	3/7/99 10:22 PM
03_08info_insert.doc	29KB	Microsoft Word Document	3/7/99 8:27 PM
03_08notes.doc	113KB	Microsoft Word Document	3/8/99 10:56 AM
03_102Final.doc	115KB	Microsoft Word Document	3/21/99 10:09 AM
03_10Answers4.doc	83KB	Microsoft Word Document	3/21/99 10:16 AM

58 object(s) 4.9MB

Note. All filenames were displayed on the computer screen in this exact fashion during each interview.