Context as a Factor in Personal Information Management Systems

Deborah K. Barreau
College of Library and Information Services, 4105 Hornbake, University of Maryland, College Park, MD 20742. E-mail: barreau@wam.umd.edu

Personal information management (PIM) systems are information systems developed by individuals for use in a work environment. Seven managers were interviewed to observe how their electronic documents were organized, stored, and retrieved. The purpose of the study was to investigate classification behavior both to identify the features of a PIM system and to suggest whether the factors which influence classification decisions in an electronic environment were consistent with the factors that Kwasnik observed for physical documents in an office. It is suggested that these behaviors may be influenced by the hardware and software environment and may impact personal as well as organizational efficiency.

Introduction

A personal information management system (PIM) is an information system developed by or created for an individual for personal use in a work environment. It includes a person's methods and rules for acquiring the information which becomes part of the system, the mechanisms for organizing and storing the information, the rules and procedures for maintaining the system, the mechanisms for retrieval, and the procedures for producing the various outputs required. Interest in PIM systems has advanced as a result of the rapid proliferation of microcomputers in the 1980s. A variety of personal information management software and hardware, such as notebook computers and pen-based input devices, have been developed to support individuals and the way they work. The industry's efforts have concentrated on making systems easier to learn and easier to use, responding to market demand for such features as graphical interfaces and multitasking environments.

Researchers into how people work and how they use systems is limited, but the studies which have been done are revealing and useful. It has been demonstrated that users frequently employ satisficing strategies when using systems, concentrating more on accomplishing job-related tasks than on learning to become more proficient with the system itself (Carroll, 1987). Furthermore, it is generally accepted that users will use what is convenient and readily accessible to them rather than attempt to find other sources which may be more appropriate but are not so easily obtained (Marchionini, in press). Work is a highly situational, context-based process which requires the accomplishment of tasks within the constraints of the working environment. The classification of work products and processes rarely fit neatly into document-specific categories such as subject and form which are common to many information storage and retrieval systems. Subject and form are important, but other factors such as frequency and date of last use, relative importance to a task, and the amount of cognitive energy required to assign a category must be considered when classifying, organizing, and retrieving information from an individual's personal environment.

Software to support PIM systems must be flexible and adaptable to the task needs and personal idiosyncrasies of the individuals who use them. To be useful to the individual, the software should help the user to integrate the processes of selecting, organizing, storing, maintaining, and retrieving information. It should be flexible enough to support the way the individual works while allowing transformation of information into formats and schemas that can be used by others. Although software environments in many places have far to go to achieve this, systems are far better and more flexible than ever before. Consequently, individuals develop unique ways of using software and of organizing and managing information within their workspaces that are highly personalized. These "personalized" tasks and strategies within a standardized business computing environment are what is described here as a PIM system.

The context in which a document is acquired or cre-
ated may have much to do with how it is classified, stored, and later retrieved. The importance of situational factors in human cognition was illustrated in studies of cues used to retrieve members of natural categories (Walker & Kintsch, 1985) and in studies of skilled memory and expertise (Ericsson & Staszewski, 1989). These studies are relevant because PIM systems are more intimate and more in tune with the way people think and work than information systems designed for general use. The purpose of this article is to examine context as a factor in PIM systems to suggest how it may influence classification decisions and ultimately retrieval. It describes a study that was undertaken to explore the factors that influence the way individuals manage information within their personal space. Interviews were conducted and analyzed to compare the patterns observed by Kwansnik in her study of the influence of context upon classification of documents in an office (Kwansnik, 1989b) with classificatory behavior of individuals in an electronic environment, and to explore the feasibility of employing the methodology in an electronic environment.

Characteristics of Personal Information Management Systems

The structure of an information system includes inputs, the individual with information needs as well as the information and entities which are part of the system; the information storage and retrieval (ISAR) system, the mechanisms for organizing, storing, maintaining, and retrieving information; and the outputs, the information or entities obtained from the system (Soergel, 1985). Figure 1 is a representation of an information system based on Soergel’s model.

**Acquisitions**

In a PIM system the user whose problems, needs, and questions are inputs to the system is usually the person who decides upon the information and the entities to be included in that system. This is not exclusively the case as any environment imposes certain constraints. Company-selected hardware or software, limitations upon access to electronic information sources, and requirements to build upon work started by someone else are some of the constraints that may impact decisions about what to include within the system. Despite constraints, much of the information acquired for a PIM system is generated by the user, or has value added by the user. Information entities may vary widely by subject and format, including formal documents that require polish, precision, and accuracy; informal notes and messages which permit irregular formats and spelling; data files; spreadsheets and tables; graphics; and program utilities, among others.

This article is concerned with inputs to information systems as they relate to the organization of information within the information space; that is, how the user defines, groups, and labels the information he or she includes within the system.

**Organization and Storage**

Mechanisms for organizing and storing information are those tools and procedures that classify, name, group, and place information in a location for later retrieval. Within a typical ISAR system, these are the access points, the controlled vocabularies, and the storage media and processes used to facilitate organization of the collection and retrieval of data relative to anticipated needs. Assigning titles, names, and subjects to items, organizing them logically, and storing them in physical locations are typical activities. Attributes such as document type, date, language, and target audience are also used to classify and distinguish information within collections.

Within a PIM system, the item attributes are often fine-grained and so personal that it is difficult to distinguish and to represent them. This difficulty is further complicated by the dynamic state of the individual’s environment as priorities are rearranged and workloads changed. Several unique attributes have been suggested for PIM systems, and it is generally recognized that multiple access points are needed. Malone found that placement of documents on a desk in an office environment often served a reminding function to the individual (Malone, 1983). This location was not based on the physical characteristics or attributes of the documents, but on the use or relevance to the person. Malone also found
that individuals would group things in a location because of the difficulty in classifying them, so items were found together which did not fit easily anywhere else, although they might have no other attribute in common. This is evidence of a satisficing strategy, of a person's avoidance of the cognitive effort required to make a decision.

Kwasnik found that the use to which a document is put or intended is often the first classification rule invoked by individuals in organizing their personal documents (Kwasnik, 1989a). Her work is described in more detail later.

**Maintenance**

Maintaining information systems is an important function, yet one often underestimated in ISAR system design. Regular system backups and routines for updating incorrect or out-of-date information are universally accepted requirements, but rules for deletion or migration vary widely from system to system. In an era of inexpensive memory and storage devices and fast processors, there has been a tendency to allow systems to grow. Individuals desiring comprehensive treatment of a subject have need to keep as much information as is relevant and feasible. Unfortunately, growth may occur with little assessment and review of the relevance or quality of the information due to the amount of time required to make such determinations. Growth can continue as long as the tools for organizing and retrieving information keep pace with the storage, and as long as system performance and personal productivity are not impaired. Decisions for what and when to delete or migrate to secondary storage are critical to efficiency, yet are frequently made at a time and under conditions dictated by circumstance rather than by rule.

**Retrieval**

Information retrieval within ISAR systems is based on user-defined queries and system access features, and the ability of the system to match the user's request to its retrieval language is one measure of performance. As with any information system, the translation of the user's query in a PIM system is not a trivial problem.

The process of retrieval in a PIM system depends greatly upon the situation or event, the information need, which initiates the search. Formal ISAR systems rely upon controlled vocabularies, structured files, and elaborate filing mechanisms to facilitate use. Within a PIM system, each person's vocabulary is itself "controlled" since it is unique for that person. It takes discipline to settle upon a subset of one's vocabulary for use, particularly when the effort required to do so interferes with the accomplishment of a task. On the other hand, there are retrieval cues available to persons in their own systems which formal systems lack. Dates, file size, and context clues from the user's situation at the time the document was created or last used are among the cues which can combine to jog a person's memory to locate an item or to retrieve a fact. Even the acts of classifying, naming, and storing a document to disk occur in the context of other work and events that may themselves become an aid to retrieval later.

**Output**

One of the most important features in the structure of information systems is output. Users come to the system to resolve a need or answer a question, and the system's capability to satisfy the request is greatly dependent upon the quality of output. For PIM systems, output must be criterion-based, subject to the needs and objectives of the individual rather than normalized to generic tasks for multiple users as in general ISAR systems. Flexibility and variability of output become important criteria in evaluating PIM systems.

In summary, one factor that defines PIM systems and that sets them apart from general ISAR systems is the unique, situational aspect of the working environment. The context in which a person works, including the nature of the task, the subject matter, the available tools, the intended audience for the work, and the intended use for the information are factors which are in play at the time the item is acquired, created, classified, stored, or retrieved. Successful PIM systems, and software used to implement these systems, should accommodate situational factors to facilitate system use.

**Context and Classification: The Kwasnik Study**

Context is the situation in which an event occurs. Context includes all aspects of a person's experience, and it has long been recognized as a factor in human behavior. We are often unable to function well when we are tired, or if environmental conditions are extreme (the room is too hot or too cold, for example), or if there are distractions. In a corporation, a manager with years of experience in the organizational culture has a perspective on what information is important for the job that differs from that of a young management trainee. Every individual is unique, possessing experiences which impact his or her decisions and reactions to new experiences, and these experiences become the basis for individual growth and learning.

People tend to analyze and to classify new objects or experiences in life by comparing information about these new objects and experiences with those which are familiar. This "reasoning by analogy" is the context theory of classification (Kwasnik, 1989b; Medin & Schaffer, 1978). Prototype theory is an extension of classification theory, implying that exemplar experiences lead to prototypes on the basis of a category's central tendency. According to the theory, classifications are based on perceptions of how well items fit the individual's idea of the
prototype. Again, each person’s experiences are unique and, even when people agree on a prototype, they may not agree upon the characteristics which define it or whether a particular object is close to the prototype or far away.

Malone explored how people organize things in the context of their offices. He interviewed ten people from various job categories to explore the patterns in their organizing behavior and the implications for office information systems (Malone, 1983). He found that individuals tend to organize information in their offices into “files” and “piles.” Files are well-organized, often labeled stacks or folders whereas piles contain miscellaneous documents that have no apparent organization or labeling. Malone discovered that the location of the files and piles on the subjects’ desks was an indication of their importance, serving a reminding function of work which needed to be done. He also found that the subjects’ descriptions of the information probes used to retrieve a document did not correspond to their descriptions of their organizational schemes in about two thirds of the cases. Malone concluded that automated systems could resolve many of the problems in these workspaces by supporting multidimensional classification and by providing devices which would serve the reminding function such as varying the size or color of icons based upon the importance of the document or having an item surrogate appear on the screen periodically.

Malone’s investigation raises questions concerning the process by which people organize and retrieve information within their workspace, and Kwasnik raises some of the same questions in her work. She investigated the influence of context upon the processes employed by individuals in organizing and classifying their own documents in their own space. Kwasnik interviewed eight researchers from various academic disciplines and asked them to give her a “guided tour” of the documents in their offices and work areas. In a second visit, subjects were asked to sort a day’s mail and to describe each piece aloud and explain what they would normally do with it. In a third visit to four of the subjects, she sorted a day’s mail for each based upon the “rules” identified in the second visit. These interviews were taped, and interviews from the first and second visits were also transcribed. She analyzed the transcripts to identify the categories of documents identified by users and the dimensions used in making these classifications. Head nouns were selected to determine how individuals name and classify documents they deal with, and to identify the dimensions used to organize them. (A software program, Qualog, a data analysis tool developed at Syracuse University, was used to facilitate the analysis.) Two examples follow:

“on the top shelf are books that are very seldom used”
“correspondence I must deal with immediately goes into my briefcase” (Kwasnik, 1992b, p. 4)

In the above examples, each decision invoked the dimensions of LOCATION (top shelf, briefcase), FORM (books, correspondence), TIME (very seldom, immediately), and USE (used, deal with). She coded and inventoried all head nouns mentioned by the subjects in the interviews. Once this task was completed, she analyzed the dimensions and organized them into categories and subdivisions (Fig. 2).

Kwasnik’s analysis produced a frequency distribution of the incidence of each dimension described in Figure 2 among the categorizations of the head nouns and general nouns identified. She defined general nouns to be those which identify documents without mentioning form; things, stuff, and items are examples. She also created frequency distributions of the occurrence of dimensions in pairs, when two dimensions were used in the classifi-
cation, and in triplets, when three dimensions were used, to observe patterns. Of the dimensions identified, the five occurring most frequently were USE, TOPIC, TIME, CIRCUMSTANCE, and FORM. When two dimensions were used, TOPIC and USE were most frequently paired, and when three dimensions were invoked, CIRCUMSTANCE-TIME-USE were used together most often. Situational attributes were invoked by her subjects more often than any category except form.

Kwasnik's work supports the notion that traditional systems which rely too heavily on document attributes are insufficient for an individual's requirements for organizing, managing, and retrieving information for personal use. Furthermore, if context is a key factor in the organization of materials within a person's physical space, it follows that context is just as important, or even more important, in how those materials are used in an electronic environment to do work. Certainly, it would be interesting to investigate classification behaviors of individuals in an electronic environment to see if the dimensions identified by Kwasnik tend to hold up. The goal of such an investigation is to suggest whether people organize information in their electronic space by similar dimensions and to explore the implications for people who work in environments where documents must be shared. The next section describes research performed with these goals in mind.

Methodology

A study was undertaken to investigate the factors that influence classification decisions by individuals in their electronic environments. Seven individuals, all managers, were interviewed and were asked to provide a guided tour of their electronic directories both to identify the documents in their workspace and to isolate factors affecting their decisions to include the documents within their systems. Managers were chosen because of the variety and quantity of information they must deal with on a daily basis. The interviews were analyzed to identify the dimensions used in organizing and retrieving information in an electronic environment.

Subjects

Seven managers were interviewed, including three males and four females, considered to be experts in their respective domains, but only two would classify themselves as experts with computers. Four of the managers are employed on the same project and work in the same computing environment, a Novell local area network (LAN). These individuals work in a large information management company providing services to government and commercial clients. One is a financial manager, one oversees document delivery functions, and two manage reference and referral services. Three of these four have computers without hard drives and are therefore restricted to the storage available on 3.5" diskettes and on the public and private directories of the file server, but the fourth has a 120-megabyte hard drive as well. A fifth subject works in the same company in a different department and supervises a group of graphic artists. She uses a Macintosh computer on an AppleTalk network which has access to a Novell LAN via MacLink Plus. This group supports a variety of projects, including the project of the four managers already mentioned.

Two of the managers interviewed were located at different sites. One was a project manager in a research department for a government agency, and many of his responsibilities are computer-related. The computing environment included a powerful microcomputer with two hard drives, operating OS/2, a multitasking operating system. This was linked to a Banyan network. The final subject works in a networked environment as well, running Microsoft Windows as the interface to his own hard drive, to a local area network, and to an IBM mainframe. This subject is a project manager/research scientist at a major corporation providing services to the government. His current project is designing a user interface for a system to be used by a federal agency.

Procedures

The managers were told in advance of the interview not to clean-up or reorganize their directories, but to leave them as usual. The interview followed the loose structure of a guided interview. Subjects were encouraged to describe freely the documents and information within their workspace, but questions were asked to elicit information about the typical functions of personal information management systems. For example, subjects typically were asked questions about where documents in their workspace originated, how they came to be where they are, when and under what circumstances they might be moved or deleted, and how they would go about finding the document when it was needed. The interview was concluded when the subject completed the tour.

The interviews were taped, transcribed, and were analyzed within the framework of a PIM system, that is, the rules and procedures applied by the individuals in acquiring, organizing, maintaining, and retrieving documents within their personal workspace. The interviewer also made brief notes during the sessions. Although none of the subjects used formal PIM system software, the data suggest that rules and procedures for managing information were present. The transcripts were analyzed to identify the documents used by individuals in their electronic workspace as well as to identify the dimensions and rules applied in classifying them. Identification of the categories of documents and the general nouns used was accomplished with ZYINDEX, commercial text retrieval software. Each occurrence of a term was analyzed in context to determine which if any of the di-
Dimensions identified by Kwasnik were applied by the individuals in their classification descriptions of the items in their electronic space. The dimensions identified were coded with the classification descriptions and were written to a file. The document categories identified that are unique to each of the individuals interviewed, and a list of the general nouns used most frequently by all subjects appear in Figure 3.

Once the document categories and rules were identified and the dimensions coded, the rules were further analyzed for their application in acquisition, organization, maintenance, and retrieval of information for PIM systems. These are discussed briefly in the sections which follow.

**Acquisitions.** The following categories of documents were specified by the subjects when asked to describe the information included in their workspace. The managers employed descriptions consistent with the dimensions identified by Kwasnik and, similar to Kwasnik's findings, situational attributes were used most often to describe documents acquired.

(1) Things from a public directory, or in a public directory, were used most often to describe documents acquired.

![Table of Unique Document Identifiers or "Head Nouns" by Individual Interviewed](image)

<table>
<thead>
<tr>
<th>Subject 1</th>
<th>Subject 2</th>
<th>Subject 3</th>
<th>Subject 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>referrals</td>
<td>window</td>
<td>correspondence</td>
<td>menu</td>
</tr>
<tr>
<td>manual</td>
<td>pun</td>
<td>modification</td>
<td>server</td>
</tr>
<tr>
<td>appraisal</td>
<td>notes</td>
<td>plan</td>
<td>newsletter</td>
</tr>
<tr>
<td>speech</td>
<td>record</td>
<td>task</td>
<td>inquiries</td>
</tr>
<tr>
<td>letter</td>
<td>joke</td>
<td>billing</td>
<td>news</td>
</tr>
<tr>
<td>decision</td>
<td>decoration</td>
<td>expenses</td>
<td>paper</td>
</tr>
<tr>
<td>recommendation</td>
<td>article</td>
<td>appropriations</td>
<td>insertions</td>
</tr>
<tr>
<td>guide</td>
<td>overview</td>
<td>draft</td>
<td>specs</td>
</tr>
<tr>
<td>policies</td>
<td>executables</td>
<td>narrative</td>
<td>briefing</td>
</tr>
<tr>
<td>agreement</td>
<td>fonts</td>
<td>bid</td>
<td>demo</td>
</tr>
<tr>
<td>search</td>
<td>talk</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bits and pieces</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject 5</th>
<th>Subject 6</th>
<th>Subject 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>costing</td>
<td>mailer</td>
<td>labels</td>
</tr>
<tr>
<td>goals</td>
<td>logo</td>
<td>requests</td>
</tr>
<tr>
<td>hotsheet</td>
<td>composite</td>
<td>analysis</td>
</tr>
<tr>
<td>thesaurus</td>
<td>jobs</td>
<td>evaluation</td>
</tr>
<tr>
<td>directions</td>
<td>archive</td>
<td>attachment</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Most Commonly Used General Nouns (used by all subjects)**

<table>
<thead>
<tr>
<th>Noun</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>thing, things</td>
<td>208</td>
</tr>
<tr>
<td>file, files</td>
<td>136</td>
</tr>
<tr>
<td>disk, diskette</td>
<td>83</td>
</tr>
<tr>
<td>stuff</td>
<td>74 (not used by subject 7)</td>
</tr>
<tr>
<td>directory, directories</td>
<td>74</td>
</tr>
<tr>
<td>network or LAN</td>
<td>43</td>
</tr>
<tr>
<td>report</td>
<td>37</td>
</tr>
<tr>
<td>information</td>
<td>27</td>
</tr>
<tr>
<td>copy</td>
<td>21 (not used by subject 7)</td>
</tr>
<tr>
<td>project</td>
<td>21 (not used by subject 7)</td>
</tr>
<tr>
<td>mail, email</td>
<td>20 (not used by subject 1)</td>
</tr>
<tr>
<td>memo</td>
<td>20 (not used by subject 4)</td>
</tr>
<tr>
<td>message</td>
<td>16</td>
</tr>
</tbody>
</table>

FIG. 3. A total of 136 unique nouns or general nouns were used by the subjects to describe the items in their workspace.
Electronic mail messages (IMPORTANT, CIRCUMSTANCE, LOCATE, FORM, ACCESS, OWNERSHIP). All subjects included items from shared space within their collections. Subjects distinguished between items that were copied from a public directory and those that were used and maintained in the public directory. In the former case, the public directory is the source for a document the subject has acquired. In the latter, the public directory is a resource for shared files such as a referral database and an accessions list used for retrieval of factual information.

Software (FORM, TIME, USE, CIRCUMSTANCE). All subjects interviewed referred to the software as a tool to accomplish work, such as "XYWrite, the word processor I use," it is clear that software is used as a dimension in classification as well for the files created with the software; that is,"Lotus files," "SAS logs," and "Harvard Graphics diskettes."

Things created by the subject which are used regularly or currently (LOCATE, USE, CIRCUMSTANCE, TIME, FORM, RELATED TO ME, TOPIC). One of the most frequently mentioned descriptors of documents within the individual's electronic environment was frequency of use. Subjects described files that were used "most days" and things they had been working on "for a couple of weeks."

Things on a relevant topic (TOPIC, FORM). One subject described his collection by saying that 75% of his files are budget-related. Others described information relevant to their projects. In the electronic environment, topic was rarely mentioned without reference to form.

Something that someone else provided (OWNERSHIP OF THE DOCUMENT, SOURCE, RELATED TO ME). Documents sent to or given to the subjects included reports for review, information to be converted into some other format or to be used in some other product, and files left by a predecessor. "George's disk," the "budget from Mike," and "someone has sent me his progress report" are examples.

A copy of something the subject created for someone else (CIRCUMSTANCE, USE, OWNERSHIP OF THE DOCUMENT, USE, VALUE). A database search that was done for someone else was kept because the subject thought it might be useful. Something "special" created for and at the request of the project monitor was kept for reference should additional work be required.

Something created for a special reason (CIRCUMSTANCE, USE). Policies and procedures were created by one subject before she went on maternity leave to help others to keep track of responsibilities in her absence. When she returned, she found it useful to maintain the information.

Things that may prevent more work later on, or that might be needed again (TIME, CIRCUMSTANCE, ACCESS, DISPOSITION). Subjects often referred to things which were kept just in case they might be needed again. There were two examples of individuals describing a situation where a document was kept to avoid work in the future. The first was a case where multiple copies of a document were kept as a precaution against loss, because it was something that is issued periodically and "has to go out quickly, be accurate." Another example was a logo, created by scanning, which required a lot of time to clean-up for use in a publication. This was saved to avoid having to redo the work should it ever be needed again.

Personal stuff (USE, FORM, CIRCUMSTANCE, RELATED TO ME). Only one subject admitted that there were personal documents in her collection. She mentioned specifically instructions for how to feed her cats, and her fiance's resume. Other subjects mentioned receiving personal messages in electronic mail.

Classification and Organization. The rules for classifying and organizing information in the subject's workspace include the dimensions described below. Often, multiple dimensions were described in a classification event, and the following list is inclusive of all dimensions applied, whether used singly or in conjunction with something else. FORM, LOCATE, TOPIC, CIRCUMSTANCE, and USE/PURPOSE were mentioned most often in classification decisions.

Things used regularly, or used currently (USE, LOCATE, CIRCUMSTANCE, TIME). All the subjects classified some items based upon their patterns of use. Examples include "my working stuff," "things we use all the time," "the main things I'm involved in," and rarely used documents which are stored "in there, out of the way."

Topical information (TOPIC, FORM). Every subject mentioned topic as a dimension in classification. Examples include "construction stuff," "OS/2 stuff," and things that "relate to AIDS." Topic was usually mentioned with other dimensions such as location and form. Kwasnik considered titles, the names of documents, as a feature of the TOPIC dimension, but in an electronic environment, the file
System Maintenance. System maintenance is defined here as modifying, moving, or deleting documents from the PIM system. More than one dimension was involved in most maintenance decisions, and Kwansnik's dimension of DISPOSITION (to keep, discard, change, locate or postpone decisions) is integral to all maintenance tasks and should be understood as applicable in each of the following rules.

1. Documents are removed once they exist in another format (NEED, USE, FORM). Some documents were removed once they were converted to another software format. In other cases, documents were created by the subject for use in some larger publication and the pieces remained in the subject's environment until the publication was done. Sometimes documents were kept only until a hard copy was obtained.

2. Documents are moved or deleted because they are finished or because they will not be needed again (NEED, USE, ACCESS, TIME, CIRCUMSTANCE). Closely related to the above rule, these are documents that were created for a specific purpose, and once that purpose was fulfilled there was no need for the document. Sometimes items were simply moved to less prominent locations "when I'm finished with it here and I may need it for something else."

3. Documents are moved or deleted based upon the time available to do this (TIME, CIRCUMSTANCE, USE, NEED). Most subjects admitted to irregular system maintenance. Most relied upon network managers to back-up their data, leaving hard drives and diskettes vulnerable. Maintenance was reported to occur "when I'm not doing anything." "I'll just be looking at them," or "in a random sort of way." During the seven interviews, each subject deleted at least one file or message. Three of the subjects stated that they removed files when they were told to by their LAN administrators.

4. Documents are deleted or moved based upon the age of the items (TIME, DOCUMENT ATTRIBUTE). During the interviews, two of the subjects deleted items because of the age of the documents. In one
Documents are deleted or moved based upon size factors such as the amount of available space, the number of files in a directory, and file size (ROOM/SPACE, PHYSICAL ATTRIBUTE, CIRCUMSTANCE). In some cases a rule of thumb was applied, such as the manager who tried to limit his Lotus directory to "30 to 35 files." In other instances it was more an act of desperation "in order to get some room on my hard drive." As might be expected, subjects with the least amount of space available performed document deletion or transfer more frequently than those who had additional media at their disposal. At least two subjects indicated they were getting additional hard drives to combat the space problem.

Documents are deleted or moved based upon the importance of the information (VALUE, CIRCUMSTANCE, FORM, TOPIC). One subject had very elaborate back-up schemes for his most important documents. Although one manager said that she keeps everything, it was clear that moving copies to storage as a safeguard against losing it was left to network administrators. Several subjects indicated that they backed-up important information, but not one person interviewed performed regular back-ups of all their files. At least one person was aware that there were no back-ups of certain important information beyond the LAN tapes, and that it would be difficult to retrieve anything from the tape back-ups that was more than a few weeks old.

Documents are maintained or updated when new information is received (CIRCUMSTANCE, USE). Most subjects used some documents which required regular modification. These items were always organized so the user could access them easily, but the location was different for each subject. Spreadsheets, databases, referral lists, and reports are among the categories of items routinely maintained.

Documents are deleted or removed because the contents are unknown to the user, or because the user does not remember the document (COGNITIVE STATE). Almost every subject had items in a directory which he or she could not identify. In some instances, these were macros or software-generated set-up files. In other instances, they were files that were left by a predecessor or files given to the subject by someone else, and due to lack of time or interest nothing had been done to identify and classify the information in a way that might be meaningful. Some subjects found items they had created in places they did not expect to find them, and could not explain or remember their reasons for putting them there.

Information Retrieval. One question asked of each subject in the interview was how he or she would go about finding a document. Each subject employed strategies for organizing information with the ease of future retrieval in mind, but many of the strategies employed in retrieving information not only took advantage of that organization, but compensated for its limitations. In a physical space, there are many simultaneous physical cues present to assist in retrieval, but in a PIM system, the individual must compensate by developing information seeking strategies which maximize the situational factors and system-defined features available. ORDER/Scheme and FORM were the dimensions cited most often in retrieval as illustrated below.

(1) Location (ORDER/Scheme, ACCESS, USE, CIRCUMSTANCE, LOCATE). One of the conventions used frequently in retrieving documents was by location. Software applications used most often were accessible from the root directory of hard drives or in prominent places on the menus of networks. Subjects kept "all my Harvard Graphics files on a disk" or turned to shared directories on the network where the referral lists, macros, or logos were always found.

(2) Names or titles (OWNERSHIP, RELATED TO ME, AUTHOR, SOURCE, ORDER/Scheme, VALUE, DOCUMENT ATTRIBUTE, TOPIC, CIRCUMSTANCE). Document names may be names that serve as a surrogate title or alternate title to fit the constraints of the 8 to 11 characters available in the MS-DOS environment. One subject said he tried to create meaningful labels in naming files so that they would trigger his memory when he went to look for something.

A typical naming convention used was to attach meaningful suffixes or file extensions to make file names more distinguishable. Examples include suffixes to group items together topically, suffixes to group them by use, and suffixes to make a distinction among documents with a similar prefix. Mnemonic devices such as file names that hint at what the item does or is were also used. An example of the former is MOVESOUP, a batch file devised by one person to move electronic-mail messages received by someone named Campbell to a subdirectory. Icons were used creatively in the OS/2 and Macintosh environments as decoration or sometimes as puns for the directory contents or for the action performed by an executable file. In the Macintosh environment which allows naming conventions to exceed the 8 to 11 character maximum of the DOS world, it was common to find full and more detailed titles.

Naming files after people was a convention employed by all of those interviewed, and included situations where the item was created for the person, produced by the person, was about the person, or concerned an area which fell under the person's responsibilities. One subject named his contract files after the last name of the contracts officer so they would be grouped in alphabetic proximity of each other for retrieval.

(3) By reordering things (DATE, USE, DOCUMENT ATTRIBUTES, ORDER/Scheme). Subjects used a variety of methods of grouping and arranging information to expedite retrieval. In some cases, they
used "arrange" functions within the software or operating system to order items by date or by title and then refer to dimensions such as file size or extension to locate the desired item. The natural order or scheme provided by Microsoft Word is "last in, first out" and the subject who worked in this environment retrieved the things he used most frequently and most recently without using any other strategy. One subject made use of the WordPerfect "find" feature to retrieve files of like extensions. Yet another subject prepared his own utilities and tools to help him navigate through his extensive workspace.

Subjects were not observed using utilities such as "whereis" or the Macintosh "find" facility which were available within their environments. It is possible that they were unaware these utilities exist, although they may have overlooked them at the time, or may have bypassed them intentionally, preferring the ordering schemes described here to command-driven find features.

(4) Date (TIME, USE). Date was a dimension used by several subjects in retrieval. Usually, it served as a discriminating factor when several candidate files had been identified as the possible target. However, there were instances where the files were displayed in date order to make it easy for the subjects to find what they had been working on most recently. The OS/2 environment offered the greatest flexibility, allowing the user to see the date the file was last used as well as the date created.

(5) Browsing (USE, ORDER/SCHEME, DOCUMENT ATTRIBUTES, ACCESS, CIRCUMSTANCE). Browsing was a strategy used frequently to retrieve files. It was the observed method of choice among the subjects when looking for old files, but it was used just as frequently to retrieve documents from a current list.

Browsing strategies varied and the activity served several purposes. Browsing was used to reduce cognitive load by perusing file lists to facilitate document recognition. This required less effort than remembering the file name and commanding the system to retrieve it. One subject entered her word processor during the interview and immediately produced a listing of the files to find a document she was working on that morning, saying aloud, "Now what did I call that file?" Browsing usually involved multiple dimensions, with date and file size often as important as the name in locating the desired item.

Most often the subjects used some kind of probe to enter the system for browsing. This probe was usually a partial filename (prefix or suffix), but sometimes it was a date or a location (a subdirectory or diskette). Subjects also demonstrated instances of browsing without a probe, broadly searching for the item needed, perhaps knowing or not knowing any part of the name, or even if what they were looking for would be there. This is the most difficult type of browsing for specific data retrieval and it requires a high level of concentration.

(6) Hard copy aids (ACCESS, ORDER/SCHEME, CIRCUMSTANCE). One subject worked in a department which used a hard copy directory and an elaborate logging system to facilitate retrieval from a vast archive of information stored on diskette.

(7) Use (USE, PHYSICAL ATTRIBUTES, FORM). Use is a powerful dimension in retrieval from PIM systems. Subjects in this study were observed often searching for files by accessing the software used to create them. These subjects relied upon the software to display the files from wherever the software stores them by default.

Discussion

There are parallels between the results of the Kwasnik study and this investigation into how individuals organize their electronic files. Many of the same types of documents were identified, and many of the same dimensions were used in making classification decisions. There are some interesting differences, however. Kwasnik attempted to identify and count incidents of dimensions used most often in making classificatory decisions, but this study focuses upon the dimensions used in the rules for organizing electronic documents in PIM systems to determine if similar rules apply.

The methodology used to analyze the data was helpful in discerning patterns in organizing and retrieving information in an electronic environment. The analysis posed some problems for the researcher, however. Often it was difficult to isolate the particular dimensions applied in classification of a document, determining whether the dimension applied was an ACCUMULATION or an ARRANGEMENT, for example. In these cases, the summary category (ORDER/SCHEME, etc.) was used instead of a more specific one. It was not always easy or even possible to determine with confidence that the category assigned was the appropriate one, especially when multiple dimensions were applied, as was frequently the case. Kwasnik reported similar problems, and computed a validity measure for her results by having the data double-coded. A validity measure was computed for these interviews as well, based upon a sample set of 35 observations. The coders agreed in identification of the dimensions 82% of the time. All of the observations in the sample set except one involved more than one dimension, and there was agreement on at least one of the dimensions in every case.

Another problem in analyzing the data was the difficulty in determining how to define a document in the electronic environment. Many of the files in the user's personal workspace were not documents in the traditional sense, but were complex database systems or tools created by or used by the individual in accomplishing work. These tools and utilities function sometimes as reference tools, but often function as mechanical devices to move files from one place to another, to open a new window, or to strip special characters from a file. Since these items occupy space in the environment and since
they can often be changed and enhanced by the user, the
decision was made to include them in the scope of this
analysis, but a case can be made against this. In an office,
a dictionary is a reference tool as well as a document,
but a ruler and calculator are only tools. In an electronic
environment, the distinction is less clear. A macro and a
batch file may be tools that perform functions, but they
are also documents in the sense that they can be edited,
stored, and retrieved in much the same way.

The interviewer's influence on the responses also
raises questions. Although the interviewer attempted to
use general terms (stuff, things, items) in referencing doc-
uments to allow the subject the opportunity to define his
or her documents more specifically, this did not always
work. Analysis of the interviews reveals two instances
where the interviewer first used a more specific docu-
ment label that was later picked up by the interviewee.
Some of the managers interviewed were known person-
ally to the interviewer, and as they were familiar with the
interviewer's interests and areas of responsibility, there
may have been some degree of posturing. The inter-
viewer oversees LAN activities, and these people were
careful to state they performed maintenance on their di-
terories when asked to do so.

Despite these obvious problems, it was useful to ana-
lyze the dimensions individuals employ in describing the
documents in their electronic space. It is useful in under-
standing how individuals work and in understanding
how PIM systems differ from general systems. Some of
these distinctions are described below.

Acquisitions

The personal computing environment offers many
opportunities for individuals to expand both their per-
sonal document collections and their access to external
sources. Unlike the physical offices in the Kwasnik study,
the workspace of these subjects spilled over to sources far
beyond the desktop, but few were maximizing the capa-
ibilities of these systems. Additionally, there were files in
each individual's workspace that were unidentified by
them, or that were known but never used. Much of this
is evidence of satisficing behavior, of the individual's at-
tention to traditional work-related tasks at the expense
of cognitive effort to gain proficiency with software (Car-
roll, 1987; Marchionini, in press). Only one person with
links to the Internet and to his local electronic bulletin
board regularly sought information there or saved items
found that looked interesting. Systems which invite
browsing and active user exploration are needed to help
reduce the cognitive load required to seek new informa-
tion and to acquire or access sources which will contrib-
ute to effectiveness on the job. It should be emphasized
that none of these users expressed a need for comprehen-
sive collections as their motives for acquiring informa-
tion was based upon need and relevance to their work.

Classification

Document attributes alone are insufficient to describe
and organize items within the dynamic environment of a
PIM system. The conditions in which documents are
created and used change, and these conditions impact
the way that individuals classify and organize informa-
tion. Rules that are applied for a period of time to reflect
the priorities of the moment may soon be abandoned or
forgotten; however, this study shows some consistency
among subjects in the dimensions applied to classifica-
tion events.

In an electronic environment, the quantity and com-
plexity of the work influence conventions used in classi-
fying and storing information. Some subjects inter-
viewed here used completely different rules for classi-
fying documents based upon the level of granularity
required to support the workload. An example of this
is the financial manager who organized budget data for
contracts by project name. One project was too large and
too complex to be supported by this scheme. This project
was divided into tasks which were organized either by the
name of the manager responsible for the task or by the
name of the task. The rule employed was, if the project
is too large to file by project name, then break the project
into tasks, and if the task has only one manager, file by
the manager's name, otherwise, file by the task name.
A similar case was observed for individuals who added
suffixes to further subdivide topical information when
their naming schemes left them without the flexibility to
assign similar prefixes to all items relevant to a topic.
This finer granularity is required by the individual to dis-
tinguish among the large volumes of data on the subject,
and he or she may extend the classification, or reclassify,
by such dimensions as a person's name, a date, or sub-
topic. While this is essential to the individual, it presents
problems for colleagues or successors who must interpret
the scheme to retrieve something from the space. Addi-
tionally, the individual must rely upon personal knowl-
edge of the growth and development of a project or task
to trigger his or her memory of former schemes.

In a physical office, folders can be tagged with the proj-
ec t name and with subheadings, and items can be stored
in physical proximity. Similar features are available in
an electronic environment, but the subjects in this study
did not typically store things this way. The subjects in
this study consistently used satisficing strategies in organ-
zizing the information within their workspace. Few had
ever created subdirectories. Only the two users whose
work was computer-related and the Macintosh user
showed knowledge of directory structures to organize
data. One subject continued to put files on diskette for
storage and for quick access, although he had ample free
space on a hard drive (100 of 120 megabytes unused).
Most subjects used software defaults to organize and dis-
play information and to browse data for retrieval. This
behavior is consistent with the high volume, varied, and
temporal nature of work performed by managers. Many documents are created quickly and with little concern for naming or organizing the information, including memos to staff, on-demand status reports, special requests, and other small projects that are the purview of management.

Maintenance

The subjects in this study performed maintenance upon the system when they were required to by their LAN managers, when they faced storage limitations, or when moving or removing files was a conscious effort to facilitate retrieval of the information later. Most of the subjects had some awareness of their options for storing information, but most had only limited knowledge of how to maintain their files or even considered that regular maintenance should be part of their routine. PIM system maintenance seems to be guided by a philosophy of “if it ain’t broke, don’t fix it,” rather than regularly scheduled maintenance to avoid crises.

There was evidence that at least one user experienced difficulty in modeling the system, or difficulty in accepting that things stored in a place that was beyond the person’s reach and grasp were safe. She felt more comfortable having files stored on diskettes and put in a box on her desk than trusting security to a LAN server that was distant and unseen.

One thing apparent from this study is that, as is true with hard copy document collections and most information systems, collection size is not static and the need for additional storage grows. The two individuals who already had substantial hard drives were expecting additional drives to be installed soon. Individuals who stored items on diskette sometimes reorganized files by format when the need for storage exceeded the one disk and a second was required. These collections can be expected to grow at least until budgets can no longer support purchase of more diskettes, hard drives, file servers, and other storage media, or until the users can no longer manage the information.

Retrieval

The people in this study retrieved information based upon a variety of dimensions. The strategy observed most often was browsing, usually initiated with a probe. When the software made this easy, individuals were observed doing this even when the document was something they had used recently or often. Although browsing involved additional steps, it required less effort than remembering the file name and the command to retrieve it. The use of browsing techniques suggests that multiple, situational dimensions may be used to retrieve information as well as to organize and store it, as browsing provides clues to age, context, format, topic, and other dimensions. Browsing provides the secondary benefit of allowing the individual to get a general overview of what is in the system. This is particularly important in a PIM system since it is an overview of their personal workspace.

Retrieval tools made available by the software or created by individuals were frequently employed to retrieve information. Examples of such tools include batch files to navigate through subdirectories, the find feature in WordPerfect, the arrange function available within software or operating environments which supports sorting by various characteristics, and software defaults that order the documents in some way. Few individuals indicated having trouble finding things, but the interviewer observed one subject who was unable to find a particular file he was seeking during the session.

Items were sometimes organized in prominent positions in the personal information space to serve as reminders of things that need to be done. This was found to be true among all the subjects who used electronic mail.

Conclusion

Classification behavior in a PIM system is a dynamic process characterized by satisficing strategies in selecting and organizing information, by browsing strategies for retrieval, and by the personal and situational factors surrounding the classification event. Classification behavior in PIM systems raises questions about individual productivity, organizational efficiency, and the quality of software and hardware used to support work.

Individuals have as much flexibility and autonomy with PIM systems as the hardware/software environment supports and their organizations allow. Individuals can organize their information as they choose, select naming conventions for their files, and design menus and desktops to meet their needs. Personal productivity, creativity, and job satisfaction are enhanced by these features, and organizations benefit when workers are productive and successful. However, the individualism which characterizes PIM systems may also pose problems for organizations. Organizations require some conformity in the electronic environment to support collaborative work, to reduce duplication of effort, to facilitate transfer of tasks from one person to another, and to enable persons to retrieve needed information in the absence of the individual responsible for a task.

This investigation suggests that electronic documents are not well ordered in a PIM system. There is evidence that items are stored often by system default on the available media with reliance upon software features for sorting and reordering information “on the fly.” Additionally, the observed behaviors suggest that some individuals employ organizational techniques similar to those used with paper documents and they have not adjusted their thinking to take advantage of features unique to the electronic environment. Marchionini found similar ten-
dencies among persons who were learning to use an electronic encyclopedia based upon the printed model (Marchionini, 1989). The physical clues which aid in storing and retrieving items in an office are replaced with very different personal and situational clues in an electronic system.

Individuals are inventive, creative, and flexible, and there is evidence that they are using software successfully to perform the work they need to do. What they are not doing well is exploring the full range of capabilities within their environments or maximizing the available storage. The challenge for software developers is for more powerful tools to aid the user in organizing, maintaining, and retrieving information, and for intuitive interfaces which improve the learnability and usability of these tools. It has been demonstrated often that expert users prefer to bypass menus for the power of command languages, but evidence here suggests that browsing strategies, even among experts, are preferred for retrieval tasks. This is consistent with the dynamic and situational nature of the work environment.

Future research should look into the similarities and differences between the PIM systems within an organization to determine the how these systems impact organizational efficiency. Additionally, a better understanding of the relationship between classification behavior and information-seeking behavior within a PIM environment over time is needed to understand the strategies that are most effective and to design systems that can support the way individuals work in these environments. Finally, it would be helpful to know if there are limits to the amount of information that can be managed effectively, whether limits may vary by individual, by system, and by situation, and if there are limits, how access to virtually unlimited electronic information through national and international networks will impact personal performance and satisfaction in a work environment.

Acknowledgments

The author thanks Dr. Gary Marchionini for his encouragement and guidance in developing the topic and for his advice on improving the manuscript. The author is indebted to Danuta Nitecki and Emily Marsh who assisted in the coding of data, and to the managers who were generous enough to give their time and thoughtful responses to questions.

References

Nitecki, D. Developing positive impact personal information systems. Unpublished manuscript.