

Christina M. Finneran, Ph.D. Candidate  
School of Information Studies  
Syracuse University

**Digital Libraries & Desktops:  
How Users View & Use Digital Libraries in Relation to their Own Private  
Collections**

This dissertation will study the interaction between frequent users and the digital libraries they regularly visit, specifically to gain a greater understanding of the factors which underly users' decision-making with regard to the storage of digital library items in their private information collections. An important component of this will be the development of meaningful insights into prevailing user perceptions of the digital library collection within the user work space, and how such perceptions influence and govern the treatment of the items found within the digital library.

This dissertation will contribute to the advancement of our understanding of user perceptions and utilization of digital libraries, especially as such relate to their own personal collections. The knowledge generated by this dissertation will help to shape and influence the development of more effective and more responsive digital library design in the years ahead.

**Introduction to and Importance of the Problem**

Libraries have always served as an extension of a user's private information collection; the fact, however, that the majority of resources were physical in form meant that this extension was rather limited. The physical library user could usually either borrow the item for a finite period of time or she could create an inferior duplicate of a portion of a work to keep indefinitely in her private collection.

Now, with the advent of digital libraries, users have the ability to save a duplicate item of the same quality in their private collections. Some digital libraries are adding features which allow users to develop personal profiles and to save relevant or favorite items to these profiles for easy access. The distinction between the collection within the digital library and the user's private collection is gradually becoming less clear. Personal digital libraries, which in this work will be defined as digital libraries' services aimed at "supporting the personal preferences of their users...[and] potentially idiosyncratic metadata at both the personal and the community level" (*Knowledge lost in information report of the NSF workshop on research directions for digital libraries*, 2003). Creating such personal digital libraries is somewhat controversial among those in the library science field because of the informality and ephemerality of such collections (Hill et al., 1999). Nonetheless, having users develop personal profiles is clearly beneficial to the growth of digital libraries.

Firstly, personal digital libraries have the potential to increase the number of visits to the site because of the facility with which users may store items they frequently access in their personal digital libraries rather than in their own private collections. Personal digital libraries also can help the digital library managers understand usage by tracking a particular user along with her initial and subsequent views of items. Additionally, personal digital libraries, if adopted by a multitude of users, can enable potential interesting features of the digital library. For example, the digital library could use characteristics from personal profiles to recommend potential items to like-minded users. Such enhancements would help build more of a community among the digital library users. Further, personal digital libraries can alleviate some of the concern publishers have about intellectual property abuses that can occur when users store the publishers' content on their personal computers.

What are the benefits of personal digital libraries to users? Why would they be motivated to create and use a personal profile when they can save and access items through their own private collection? Before answering these questions about personal digital libraries, it would be advantageous to understand users' storage behavior with regard to digital libraries. When and why do users choose to store useful items from a digital library in their private collection?

This dissertation seeks to understand the behavior and motivation of users who frequently visit and utilize a digital library. Specifically, it will investigate their patterns of saving items from a digital library onto a user's desktop, or other private file space. The goal will be to uncover users' needs (e.g., the need to access an item when she does not have Internet access; the need for an item to be in a project folder to function as a reminder when working on that project) which, in the aggregate, will serve to explain why and how users save and organize items in their own private collections.

Further, the dissertation will examine users' conceptual views of how the digital library fits into their collection space. Marshall (2003) notes that digital library boundaries are often unintentional and mere interpretations. She defines a boundary as, "something that tends to separate, to interpose; a boundary is a perceptible seam in the social fabric, the technological infrastructure, or a physical setting or may span all three". Collections, she notes, are boundaries that are intentional and interpreted. This study will investigate interpretations of digital libraries, interpretations of private collections, and the relationship between them.

## **Research Questions**

The following research questions guide this study:

1. How do users manage the material that they find useful from digital libraries which they frequently use?
  - Why do users store items from a digital library in their own private collection?

- In what contexts do users store items from a digital library in their own private collection?
    - How are items that are stored in private collections used? Differently than those that are not stored in private collections?
    - How are the items stored in the user's private collection conceptualized? As ephemeral, working, or archival documents?
2. How do users conceptualize the distinctions between, and the inter-relationships of, digital libraries and their private collections?
    - To what extent, and in what manner is the users' view of the digital library impacted by a fuller utilization of the personal digital library tools available within the digital library?
    - In what respects does the user differentiate the digital library from non-private collections the user has access to (e.g., physical library, general Web)?
  3. How does the way users conceptualize a digital library impact the management of items from a digital library?

## **Current Knowledge**

This research is informed by three primary literatures: the stream of research now referred to as “personal information management”, the emerging field of personal digital libraries, and conceptual aspects of the metadata literature.

### *Personal Information Management*

The research stream within Library & Information Science (LIS) and Human-Computer Interaction (HCI) referred to as "personal information management" (Bergman et al., 2004) is focused on what users do with information resources after initial discovery to enable future access and use of them. It entails studying the users' systems and strategies for storing and accessing their own private collections

Prior to the digital age, paper file systems were studied and provide some important background. Most individuals struggle with developing extensive document management systems because creating categories to structure their files is cognitively taxing (Lansdale, 1988) and an immediate benefit is rarely evident to them (Cole, 1982). Malone's (1983) noteworthy study found two basic methods that office workers use to organize their paper documents: files and piles. “Files” are named and ordered entities whereas “piles” are untitled entities that are not ordered and usually spatially located. This finding resulted in Xerox PARC's innovation of the “desktop” metaphor of early personal computers, which was adopted by Apple and Microsoft. An important function of the “desktop” was to enable users to replicate “piles”, or documents that are not filed and can be placed to serve as reminders.

When individuals do develop classification systems for their print office documents, they consider the situational factors, such as use, as well as document-specific attributes

(Kwasnik, 1991). These findings hold with electronic document filing systems within personal computers: "the way information is used is a primary determinant of how it will be organized, stored, and retrieved in the personal workspace" (Barreau & Nardi, 1995).

For print documents, users develop organization systems that included a reminder function (Malone, 1983), a feature which has not been effectively incorporated into electronic file management systems. Email has evolved into a task management system as well as a document management system (Mackay, 1988; Whittaker & Sidner, 1996). Even though it was not designed with document management capabilities, email is serving this capacity partly because it serves as a reminder and partly because it embeds contextual metadata with documents, as discussed in the metadata section below.

Jones, Bruce, and Dumais (Bruce, Jones, & Dumais, 2004a, 2004b; Jones, Bruce, & Dumais, 2001; Jones, Dumais, & Bruce, 2002) have been catalysts in the development of the "personal information management" research stream. Their research, *Keeping Found Things Found*, investigates how knowledge workers manage Web-based information for re-use. They observed and interviewed information professionals as to how Web-based information that would be accessed at a future time was stored. Rather than using the standard Web browser bookmarking function, the users used an array of techniques such as emailing themselves the URL, saving the Web page as an electronic file, and printing the Web page. The researchers developed a list of features that influenced which method the user utilized:

- Portability of information
- Number of access points
- Persistence of information
- Preservation of information in its current state
- Currency of information
- Context
- Reminding
- Ease of integration
- Communication and information sharing
- Ease of maintenance

Interestingly, in the *Keeping Found Things Found* research, Jones, Bruce & Dumais uncovered that in some cases, users prefer not to take any action to help them re-find a useful resource at a later date. These users merely "leave" the resource where it is, as they expect such resource will be readily located later if needed.

When this is considered within the context of digital libraries, which -- as opposed to the general Web -- are designed to persist and help users find resources, it is important to understand the context in which users "leave" the resource in the digital library and the contexts in which they "keep" the resource in their own private collection.

One might tend to assume that items which will always be available would be left until needed. Research in the paper world has shown that users retain an abundance of duplicative and publicly available documents. An empirical study of paper archives

among office workers found that 51% of subjects' paper documents were not unique documents but replications, that 36% of the documents were publicly available documents, and that 15% of the retained documents were unread (Whittaker & Hirschberg, 2001). Through interviews, Whittaker and Hirschberg found that these redundant local documents were kept because they enabled fast access, served as a reminding device, were of sentimental value, or were retained because of a distrust of centralized storage. Further, many redundant documents were unread and retained because the user was not able to process and evaluate them expeditiously.

Does this hold in the digital world? Will frequent digital library users also save a large percentage of redundant items? Will the same motivations hold for frequent digital library users?

Prior research has established that users do not treat all documents the same. Rather, three different types of documents have been identified based on their use: ephemeral documents, working documents, and archival documents (Barreau and Nardi, 1995). This typology has been confirmed with different computer systems (Ravasio, Schar, & Krueger, 2004). In their noteworthy monograph, *The Myth of the Paperless Office*, Sellen and Harper (2002) concur with these three document types and use a parallel of hot, warm, and cold documents for the time-sensitive industry (suppliers) they studied. This research will study if digital library users treat items differently, depending on whether they deem the items to be ephemeral, working, or archival versions.

Sellen and Harper (2002) argue that, "we are not headed toward offices that *use* less paper but rather toward offices that *keep* less paper." Can we say that in 2005 this is not only about keeping less paper but also about keeping fewer files? Will access be enough for users if they believe they will have persistent and efficient access?

### *Personal Digital Libraries*

An emerging area of research within the field of digital libraries is personal digital libraries. In the summer of 2003, a group of digital library scholars from the LIS and computer science fields were brought together to discuss the digital library research agenda for the next decade. They noted that the management of "personal digital libraries", or an individual's own repository, will be an important component because individuals will develop extensive repositories of their own scholarly or creative works and will frequently choose to store other relevant documents (*Knowledge lost in information report of the nsf workshop on research directions for digital libraries*, 2003). In their *Managing the Digital University Desktop* (n.d.) Project Tibbo and collaborators are investigating the file management practices of public university employees and developing best record management practices so that enduring value documents are archived.

A workshop, funded by the National Science Foundation & the European DELOS Network of Excellence on Digital Libraries on the personalization of digital libraries concluded that, "Personalisation is required to make an increasingly heterogenous population of digital libraries accessible to an increasingly heterogenous population of

users” (Smeaton & Callan, in press). One of the important research challenges identified was the development of more complex user models which take into account context, tasks, and situations.

The Alexandria Digital Earth Prototype (ADEPT), an extension of the Alexandria Digital Library to support course preparation and presentation, has added a personal digital library feature. An instructor can store her course material within her ADEPT personal digital library. The contents of her personal digital library can also be shared with others if desired. “The personal digital library architecture will allow collections and services to be tailored to individual practices without compromising the organization of the larger database” (Borgman et al., 2005). The researchers found the personal digital library feature important because of the unique way different users select, collect, and organize their resources.

As personal digital libraries are emerging as features of digital libraries, we need to better understand the user needs and motivations which are factors driving this emergence.

### *Metadata*

Though metadata has typically been associated with the system side of designing digital libraries, a broader definition of metadata will help inform this user-focused research.

When elaborating on the largely uninformative definition of metadata as “data about data”, Daniel & Lagoze (1997) stated:

- *There is no essential distinction between data and metadata. We can only make such a distinction in terms of a particular ‘about’ relationship. As a result, what is metadata in the context of one ‘about’ relationship may be data in another.*
- *There is no single ‘about’ relationship. There are many different and important relationships between data resources.*

Within the digital library context, metadata is not merely a surrogate of an item, but extends into connections between works and between parts or aspects of one work and other works. All of a work’s metadata will not necessarily be established before use. Contextual metadata, as it emerges as an important access point within digital libraries, will evolve with the use of the work.

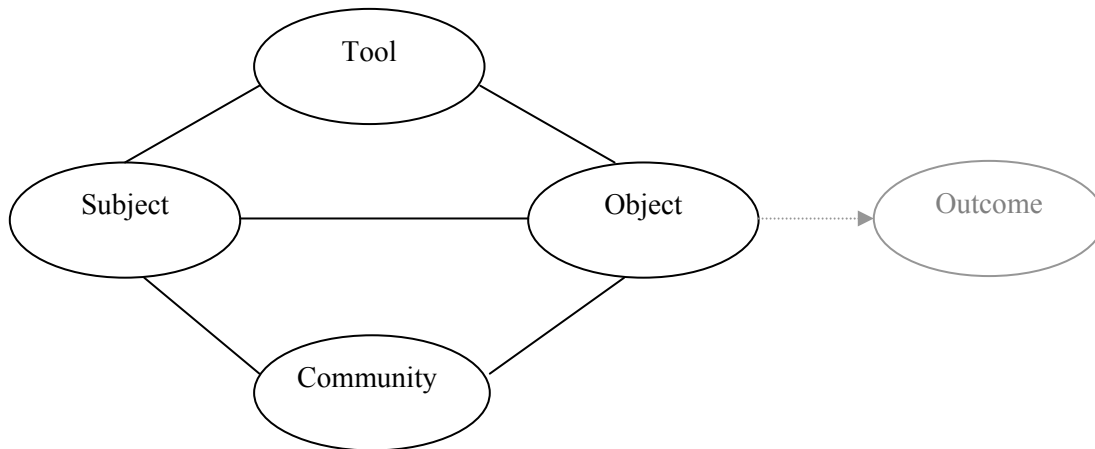
Empirical studies have shown that metadata can be personal, transient, and situation-specific (Bishop, 1999; Marshall, 1998). Email clients have become a defacto filing system of work documents partly because they can easily incorporate contextual metadata into a document (e.g., who sent it, who else received it, in relation to what project), without necessarily requiring users to develop their own classification or filing systems. Ducheneaut and Bellotti (2001) use the term "embedding" for using email as document exchange. They explain, "Exchanging documents is not a standalone activity; it is part of a wider context of exchanges aimed at accomplishing tasks. Communications form the context of document exchange; it is natural then that documents get included in them" (Ducheneaut & Bellotti, 2001). This context within the communication can be described with metadata.

While this investigation focuses on users' information storage behavior and motivation, it will also consider the metadata about particular resources that users are seeking to establish or preserve. For example, a user may want to establish metadata about the use of an item (e.g. that it might be useful for a particular project) by taking a particular action (e.g., storing it in the project folder within their file system).

### Theoretical Framework

The theoretical perspective of this study will be activity theory (Leont'ev, 1978), which emerged from the Russian psychologists Vygotsky and Leont'ev in the 1920s. In the 1990s, activity theory was adopted by members of the human-computer interaction (HCI) community to incorporate contextual factors that traditional HCI techniques like task analysis did not consider. Activity theory preserves the integrity of the, "users' own points of view, with their own descriptions of their objects" (Nardi, 1996).

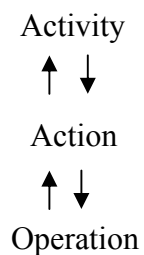
Activity theory is not a predictive theory, but a framework that "has a well-articulated conceptual apparatus and a core set of concepts that are useful for empirical analyses of HCI problems" (Nardi, 1996). The foundational idea behind the theory is that actions can not be studied outside of their context, like Suchman's (1987) work argues, and that activities are the, "minimal meaningful context for understanding individual actions" (Knutti, 1996). An activity represents an individual's motive in endeavoring to fulfill a specific objective, that is, transforming an object, which could be a person, thing, or intangible. The interaction between the subject and object is mediated by a tool and a community, or "those who share the same object" (Knutti, 1996). The resulting reciprocal relationships are shown in Figure 1.



**Figure 1. Basic structure of an activity (Knutti, 1996).**

In this research, activity theory will assist in maintaining the context of the users, exploring motive with action, and conceptually clarifying the activity the user is involved in and the various levels of tasks within an activity. One of the main principles of activity theory is the hierarchical and dynamic structuring of an activity (Figure 2).

Activity is the highest level and describes the motive or the “why” of something; the activity is the process of transforming an object into an outcome. Frequently persons will not be consciously thinking about their motive (or activity) while in the midst of work. An Action is the next level down and is usually what the person is consciously thinking about. The action involves the goals, which can be broken down into sub-goals or lower level actions. Before an action is initiated, a person establishes a model, however complete, in their mind. This phase is referred to as orientation. What has been typically called tasks can be considered actions. Below actions are operations. Operations are not consciously thought about; they are internalized actions which have become habitual. Only when the conditions change are the operations brought into the user’s conscious mind and then move to conscious actions. For example, when a user is faced with a new split keyboard, she consciously thinks about how to type.



**Figure 2. Hierarchical levels of an activity (Knutti, 1996)**

Activities will provide the structure to analyze the user’s actions. In the information seeking literature, it has been found that while users are working on one search, they may come across something useful for another search (Borgman et al., 2005; Spink, Ozmutlu, & Ozmutlu, 2002). It is likely that even while working with the same documents, users may be involved with different activities, perhaps with competing motives. For example, the user may be re-finding an article to verify a fact (for the activity of writing a paper) yet might also be involved in the activity of creating an archive of her research. These activities yield different types of actions and motives. By separating out the activities, and their associated motives, we can come to clearer understandings of user behavior.

## **Methodology**

As this research aims to learn about the users’ behavior, motivation, and conceptualization, it will use semi-structured interviews of users. In order to increase our understanding of heterogeneous users, a purposive sampling method will be used to select the participants (Babbie, 1990). First, three digital libraries that vary in type of content (e.g., journal based, multi-media) and scope of personalization features will be selected. After the digital libraries have been selected, persons who are frequent users of one of the digital libraries will be selected through purposive sampling.

The interviews will be guided by the research questions. The users first will be asked first to describe how they generally use the digital library, then asked to describe in detail the last substantive use of the digital library. The interview will focus in on such visit,



having the user walk through what she was doing, how she used the digital library, and what she did with any useful items. Next, the participant will be asked to explain the rationale behind each action she took. Finally, the user will be asked to broadly describe what she does with documents she encounters in the digital library that may be useful to her, and the situations where she may take an alternative action.

Each interview will be recorded, transcribed and content analyzed. The interview data will be analyzed at different levels. First, the data will be analyzed and coded with the activity the user is involved with, then coded for the action that the user takes and the rationale given for the action. This will be done using a content analytic tool such as Nvivo, AtlasTI, or HyperResearch.

To investigate the conceptual level, the transcript will be intellectually analyzed to elicit underlying metaphors. Metaphor analysis (Kendall & Kendall, 1993; Koch & Deetz, 1981) will uncover the particular aspects of the digital libraries the users focus on. Lakoff & Johnson (1980) explain, “In allowing us to focus on one aspect of a concept...a metaphorical concept can keep us from focusing on other aspects of the concept that are inconsistent with that metaphor.” Identifying these metaphors that exist underneath the surface will increase our understanding of different user perspectives of digital libraries.

Interviews will be conducted and analyzed until saturation occurs. All of the data will be analyzed by activity to unearth patterns of behavior among similar activities, which will inform us about contexts. Data analysis will identify patterns that occur within particular digital libraries, which will inform us about content types or enhanced features that influence behavior. The investigation will also include analysis at the user level, to see if a typology of users emerges.

### **Expected Contributions**

As digital libraries have started to venture into providing venues and means for users to create their personal digital libraries within the library, it is important that digital library design provide enlightened systems that are responsive to, and effectively serve the needs of, digital library users. This can be best accomplished by understanding the behavior of frequent digital library users. This study will contribute to digital library scholarship (disseminated through conferences, such as JCDL) by increasing our understanding of users’ behavior, motivation, and perception of digital libraries, especially in relation to the users’ own private information collections. At a more abstract level, this study will help us understand users’ conceptualization of digital libraries and how they relate to other resources they have access to (e.g., books on their shelves, files on their computer). In the aggregate, this study will provide understanding which will enable enhanced digital library design, to better serve and meet the needs of digital library users in the increasingly digital years ahead.

## References

- Babbie, E. R. (1990). *Survey research methods* (2nd ed.). Belmont, CA: Wadsworth.
- Barreau, D., & Nardi, B. A. (1995). Finding and reminding: File organization from the desktop. *SIGCHI Bulletin*, 27(3), 39-43.
- Bergman, O., Boardman, R., Gwizdka, J., & Jones, W. (2004, April 24-29). *Personal information management*. Paper presented at the CHI 2004, Vienna, Austria.
- Bishop, A. P. (1999). Document structure and digital libraries: How researchers mobilize information in journal articles. *Information Processing and Management*, 35, 255-279.
- Borgman, C. L., Smart, L. J., Millwood, K. A., Finley, J. R., Champeny, L., Gilliland, A. J., & Leazer, G. H. (2005). Comparing faculty information seeking in teaching and research: Implications for the design of digital libraries. *Journal of the American Society for Information Science and Technology*, 56(6), 636-657.
- Bruce, H., Jones, W., & Dumais, S. (2004a). Information behaviour that keeps found things found. *Information Research*, 10(1).
- Bruce, H., Jones, W., & Dumais, S. (2004b). *Keeping and re-finding information on the web: What do people do and what do they need?* Paper presented at the American Society for Information Science & Technology, Providence, RI.
- Cole, I. (1982). *Human aspects of office filing: Implications for the electronic office*. Paper presented at the Human Factors Society, Seattle, WA.
- Daniel, R., & Lagoze, C. (1997). Extending the warwick framework: From metadata containers to active digital objects. *D-Lib Magazine*(November).
- Ducheneaut, N., & Bellotti, V. (2001). E-mail as habitat. *Interactions*, 30-38.
- Hill, L. L., Janee, G., Dolin, R., Frew, J., & Larsgaard, M. (1999). Collection metadata solutions for digital library applications. *Journal of the American Society for Information Science*, 50(13), 1169-1181.
- Jones, W., Bruce, H., & Dumais, S. (2001, November 5-10). *Keeping found things found on the web*. Paper presented at the CIKM'01, Atlanta, GA.
- Jones, W., Dumais, S., & Bruce, H. (2002). *Once found, what then? A study of "keeping" behaviors in personal use of web information*. Paper presented at the American Society for Information Science & Technology, Philadelphia, PA.
- Kendall, J. E., & Kendall, K. E. (1993). Metaphors and methodologies: Living beyond the systems machine. *MIS Quarterly*, 17(2), 37-47.
- Knowledge lost in information report of the NSF workshop on research directions for digital libraries*. (2003). Chatham, MA: National Science Foundation.
- Knutti, K. (1996). Activity theory as a potential framework for human-computer interaction research. In B. A. Nardi (Ed.), *Context and consciousness: Activity theory and human-computer interaction* (pp. 17-44). Cambridge, MA: MIT Press.
- Koch, S., & Deetz, S. (1981). Metaphor analysis of social reality in organizations. *Journal of Applied Communications Research*, 9(1), 1-15.
- Kwasnik, B. H. (1991). The importance of factors that are not document attributes in the organization of personal documents. *Journal of Documentation*, 47(4), 389-398.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: The University of Chicago Press.

- Lansdale, M. (1988). The psychology of personal information management. *Applied Ergonomics*, 19(1), 55-66.
- Leont'ev, A. N. (1978). *Activity, consciousness, and personality*. Englewood Cliffs, NJ: Prentice-Hall.
- Mackay, W. (1988). More than just a communication system: Diversity in the use of electronic mail. *ACM Transactions on Office Information Systems*, 6(4), 380-397.
- Malone, T. W. (1983). How do people organize their desks? Implications for the design of office information systems. *ACM Transactions on Office Information Systems*, 1(1), 99-112.
- Managing the digital university desktop* (n.d.). Retrieved March 30, 2005, from the World Wide Web: <http://ils.unc.edu/digitaldesktop/Info/index.html>
- Marshall, C. C. (1998). *Making metadata: A study of metadata creation and use in a mixed physical-digital collection*. Paper presented at the 3rd ACM International Conference on Digital Libraries.
- Marshall, C. C. (2003). Finding the boundaries of the library without walls. In N. A. van House & A. P. Bishop & B. P. Battenfield (Eds.), *Digital library use: Social practice in design and evaluation* (pp. 43-63). Cambridge, MA: The MIT Press.
- Nardi, B. A. (1996). Some reflections on the application of activity theory. In B. A. Nardi (Ed.), *Context and consciousness: Activity theory and human-computer interaction* (pp. 235-246). Cambridge, MA: MIT Press.
- Ravasio, P., Schar, S. G., & Krueger, H. (2004). In pursuit of the desktop evolution: User problems and practices with modern desktop systems. *ACM Transactions on Computer-Human Interaction*, 11(2), 156-180.
- Sellen, A. J., & Harper, R. H. R. (2002). *The myth of the paperless office*. Cambridge, MA: The MIT Press.
- Smeaton, A. F., & Callan, J. (in press). Personalized and recommender systems in digital libraries. *International Journal of Digital Libraries*.
- Spink, A., Ozmutlu, H. C., & Ozmutlu, S. (2002). Multitasking information seeking and searching processes. *Journal of the American Society for Information Science and Technology*, 53(8), 639-652.
- Suchman, L. A. (1987). *Plans and situated actions: The problem of human machine communication*. Cambridge: Cambridge University Press.
- Whittaker, S., & Hirschberg, J. (2001). The character, value, and management of personal paper archives. *ACM Transactions on Computer-Human Interaction*, 8(2), 150-170.
- Whittaker, S., & Sidner, C. (1996). *Email overload: Exploring personal information management of email*. Paper presented at the SIGCHI, Vancouver, BC.