Chapter 1. Information and Information Seeking

For knowledge, too, is itself a power. (Francis Bacon, De Haeresibus)

Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information? (T.S. Eliot, The Rock)

Our world continues to become increasingly complex, interconnected, and dynamic--there are more people and institutions, they engage in more relationships and exchange, and the rates of change continue to grow. This is due largely to developments in technology and the importance of information to human and technical development. We live in an information society[1] where more people must manage more information, which in turn requires more technological support, which both demands and creates more information. Electronic technology and information are mutually-reinforcing phenomena and one of the key aspects of living in the information society is the growing level of interactions we have with this complex and increasingly electronic environment. The general consequences of the information society are threefold: larger volumes of information, new forms and aggregations of information, and new tools for working with information.

First, we find ourselves dealing with more information in all aspects of our lives. More of us are "knowledge workers", generating, managing, and communicating information to produce and provide goods and services for an increasingly global economy. In addition to the often noted trend toward more people managing more information in the workplace, people must go beyond the workplace to learn new skills and develop new knowledge for doing their jobs. Acquiring new knowledge is no longer done only to prepare for a career, but rather is an essential part of "knowledge work." Zuboff (p.395) has noted " To put it simply, learning is the new form of labor." Lifelong learning as part of the job has long been an important part of professional responsibilities and has spread to all venues of the labor force. Over 57 million adults participated in some form of adult education in 1990-91, of which, almost 13 million were professionals (Statistical Abstract of the United States 1993).

Not only are we required to continually seek and acquire information, but there are increasingly more sources and larger volumes available. Consider the magnitude of the following examples: In 1993 there were 11,296 newspapers and 10,857 periodicals published in the United States alone (Statistical Abstract of the United States 1993); there were 5,500 new book publishers and 136,400 new books added to the 1.5 million books in print (Books in Print, 1993); and the U.S. Government Printing Office offered over 20,000 titles and the National Technical Information Service offered 2,000,000 titles (Informing the Nation: Federal Information Dissemination in an Electronic Age). Not only are the volumes astounding, but the numbers are growing rapidly. For example, 90,300,000,000 pieces of first class mail were sent in 1991 versus 60,300,000,000 in 1980

(Statistical Abstract of the United States 1993); in 1986 there were 378,313 research articles published worldwide in science and engineering alone, versus 267,354 in 1976 (Science and Engineering Indicators-1991); there were 2490 cable television systems serving 4.5 million subscribers in 1970 and 11, 075 systems serving 57.2 million subscribers in 1992 (Statistical Abstract of the United States 1993).

These large volumes of information are organized into many collections that require secondary and tertiary indexes and directories that in turn grow in size and complexity. The growth of directories and indexes is reflected in competition among different companies that offer phone directories and bibliographic databases. The development of new and alternative organizational structures for dealing with large volumes of information in turn demands more information management skills. In our personal lives, billboards, newspapers, mail, telephone, and television serve as vehicles for incessant information assaults on our senses. To cope with these large amounts of information in our lives we develop complex personal information infrastructures--which require time and effort to build, maintain, and use. These structures include conscious and unconscious filtering and finding strategies for achieving our immediate goals and protecting ourselves from information overload. There is a tension between the goals and plans we make and the information resources necessary to achieve them; we travel a narrow road toward our goals with a sea of seductive information to distract us on one side, and a spiraling abyss of confusion and information overload on the other. Technology accelerates the rate at which we are able to travel toward our goals, but it also increases the scope and peril of the two sides.

Second, we deal with information in new forms, especially, electronic digital forms that are more abstract, more dynamic, and more malleable than printed or painted information. Much more information is becoming available in electronic form. The number of publicly available databases has grown from 400 in 1980 to 8400 in 1993; the number of database producers has grown from 221 to 3260 and the number of online services has grown from 59 to 825 during this same period (Gale Directory of Databases, 1993). Additionally, over 3500 CD-ROM titles were available in 1993 (CD-ROMs in Print 1993: An International Guide). Although much of this information is also available in paper or microfiche forms, there is a burgeoning amount of information available only in electronic form. For example, 7.5 billion packets were sent over the Internet in the month of May 1991 and this monthly volume had grown to 13.4 billion packets by February of 1992 (Merit/NSFNET Link Letter 5(1), p.16).

Information in electronic digital form is both enabling and complicating. On the one hand, electronic digital information is more accessible--available from anywhere in the world with a few computer keystrokes or mouseclicks. On the other hand it is less accessible because it is not directly perceivable to humans unaided by technology. We are dependent on machines to express this information in forms that we can perceive. Electronic digital information is manipulable--it allows us to apply the computational

power of computers to systematically aggregate, classify, compare, change, and transmit information. Electronic digital information forms allow copies to be made perfectly and recursively, unlike analog or physical forms that degrade over generations of copies. Electronic digital information is simple because it is fully expressed by only two elements (bits), however, it is complex because many levels of coding schemes must be used to map the enormous variety of structure and meaning in the world into binary form. The many sets of codings necessary for humans to "make sense" out of digital information allows the same digital code to be represented many ways; for example, the set of bits 1001101 can be expressed on a display as an upper case "M" or as a set of black and white pixels in a larger image, or as half of a note value for a compact disc recording. Standard coding schemes (e.g., American Standard Code for Information Interchange-ASCII, Tag Image File Format-TIFF, Digital Alternative Representation of Musical Scores-DARMS), facilitate communication and exchange of information, but the many possibilities support a kind of information alchemy, where words, numbers, images, and sounds can be interchanged--for better or worse. Given the sound and graphic editing tools available, it is no longer possible to firmly believe that digitally recorded sounds or images represent reality. The implication for humans is that additional levels of learning and cognitive effort are necessary to use, interpret and validate information based on electronic digital expressions.

Third, we find ourselves using new tools to manage information--tools that we must learn to use, pay for, and maintain. The primary tool of the information society is the computer[2]. Microprocessors are used to improve the performance of other technologies and computers are increasingly used by individuals to control and integrate other kinds of information technology (e.g.,TV, radio, telephones)[3]. Computer literacy has become a component in primary and secondary school curricula in all industrialized countries, and billions of dollars a year are spent on training and upgrading workers' computer skills. As more computing technology is developed, more new learning and retraining will be needed, placing demands on our time and financial resources. The computer industry accounts for an increasingly large share of the gross national product of the industrialized countries and the massive personal computer market has driven the invention of new software tools that fit the needs of a great variety of users. The need to produce products that can be used by the general population has in turn driven advances in human-computer interface research. Although good progress has been made in making computers easier to use, the evolution of hardware and software, and the rapid pace of information creation and manipulation will require that significant material and intellectual resources be devoted to acquiring, learning to use, applying, and maintaining electronic tools for the foreseeable future. At the very least, it is obvious that more and more of our time and financial resources are devoted to using computers, and we will become more dependent on them in the future.

An important aspect of these effects is that more of our professional and personal lives are spent interacting with complex systems. Interactions with other people are

physiologically natural, psychologically necessary, and culturally expected. These interactions became increasingly mediated as communications technology developed, and computer technology is another step toward intermediated personal communication. This trend is illustrated by use of electronic mail which had 9.2 million users in 1992, and is projected to have 15 million in 1993 and 38 million users by 1995 (Reinhardt, 1993). Individuals also interact with a variety of institutional systems such as, government agencies, businesses, and other organizations. These interactions were traditionally mediated by other people, but information technology is finding increasing application. Consider, for example, automatic teller machines, phone menu systems, and information kiosks at shopping malls or museums. As more of these systems are connected together, many of our interactions with the institutions of civilization and with other people will take place through electronic workstations rather than through personal contact. Working from one's workstation conserves time and resources and the computational power facilitates execution of multiple tasks concurrently. The implications of mediated communication, high rates of exchange, parallel processing of tasks are considered under the topic of interactivity.

Information is a valuable resource in an information society, thus acquiring and using information are critical activities. The process known as information seeking is therefore becoming more fundamental and strategic for intelligent citizenship. Additionally, the information-seeking process is increasingly dependent on electronic technology. This book examines the physical, cognitive, and affective consequences of electronic environments on the increasingly important process of information seeking and provides a framework for designing systems that support information seeking. It takes the point of view that information seeking is dependent on the interactions among information seekers and other people and systems for representing information. It argues that highly interactive electronic information systems are causing incremental changes in how we seek, acquire, and use information. The high volume and diverse forms of information demand better tools which in turn change our behaviors, expectations, and attitudes. At this stage in the evolution of the information society, we argue for designs that place users in control of highly interactive environments that focus on content rather than on forms and tools.

WHAT IS INFORMATION?

The word "information" will be used to refer to several different concepts in this book. Buckland (1991) has distinguished information-as-process (the communication act), information-as-knowledge (an increase or reduction in uncertainty), and information-as-thing (the objects that may impart information). In this vein, he also distinguishes the actual knowledge in a human mind (what one knows) from the artifacts of the world that represent knowledge. Most generally, *information* is anything that has the potential to change a person's knowledge. This sense, after Belkin (1978), admits reflection on one’s memory traces, the objects that convey information, and the ideas

and knowledge contained in other minds. Thus, information will be used in a general manner that includes objects in the world, what is transferred from people or objects to a person's cognitive system, and as the components of internal knowledge in our minds. For the purpose of information seeking, people seek to change a state of their knowledge and seek physical representations (e.g., ink on paper, sound waves, electronically charged phosphorus, etc.) that represent abstractions (e.g., words, numbers, images, concepts, melodies, etc.) that have the potential to cause this change.

Because there are many manifestations of information-as-object, there are many terms that will be used to describe those objects. Terms such as bit, data, record, text fragment, graphic, document, utterance, database, book, and library are all used to label particular information units. Although these terms are typically associated with different media or information systems, the terms "*document*" and "*information object*" will be used in a general way to represent information-as-object. Thus, documents may be considered as a single numeric value, a database record, a distinct image, or a video segment, as well as the more typical textual collection of words related to a topic.

INFORMATION SEEKING

Much of human existence is characterized by the notion of search; we seek and pursue material objects such as food or shelter, sensual experiences such as adventure or ceremony, and ethereal objects such as knowledge or justice. We are concerned here with the search for information which we will call information seeking. *Information seeking* is a process in which humans purposefully engage to change their state of knowledge. The term *search* will be used to mean the behavioral manifestation of humans engaged in information seeking and will also be used to describe the actions taken by computers to match and display information objects. The term "information seeking" is preferred to "information retrieval" because it is more human-oriented and open-ended. Retrieval implies that the object must have been "known" at some point; most often, whoever "knew" it organized it for later "knowing" by themselves or someone else. Seeking connotes the process of acquiring knowledge; it is more problem-oriented; the solution may or may not be found. For example, seeking spiritual enlightenment makes sense, but retrieving enlightenment does not. Retrieval is applicable to database management and most applied problems, but seeking is closer to question-answering or learning.

Information seeking is a fundamental human process closely related to learning and problem solving. Nature has evolved tools and methods to support information seeking, resulting in physiological and psychological abilities that are well-suited to information seeking. Our perceptual organs gather massive streams of environmental data, our muscles aim these organs and carry us closer to the objects of search, and our cognitive and emotive engines direct muscles and organs and process the incoming data. Our cognitive processors adopt various organizational structures and systematic

strategies for filtering, comparing, and storing information in a variety of media. Our emotive selves derive stimulation and pleasure from seeking and integrating information. Information seeking is thus a natural and necessary mechanism of human existence.

The information-seeking processes needed to survive and prosper have become increasingly complex as social organizations have developed. The ability to locate and apply information is an important component of what it means to be literate. Just as nature has evolved physiological and psychological tools and methods to support information seeking, so culture has evolved tools and methods to support information seeking. Information processing technologies from the abacus to the zoetrope aid us in generating, manipulating, and representing information. As social and economic organizations have become more complex, so has the information necessary to work in these organizations, and this has led to new, more powerful technologies for managing information. Today, generation, storage, and communication of information are inextricably linked with technology--it is virtually impossible to conduct business in many markets today without technology to assist us in managing the generation, storage and flow of information. Likewise, an enormous amount of information is necessary to select from entertainment options or practice good consumer decision making. Thus, one of the key changes in the information society is that information seeking has become a fundamental skill for larger portions of the population--more people must regularly manage more information to survive and prosper and they must use an expanding array of technologies to do so.

Information seeking, like learning, is a fundamental and high level cognitive process. Information seeking is often part of learning or problem solving, but it is also distinct. Information acquired during learning is purposefully stored so that it can be recalled and used at a later time, however, information acquired as a result of information seeking may be useful for a specific task and then discarded. Intermediate or temporally-relevant information often should be discarded so that it does not take up storage space or complicate the organization of stored information and subsequently interfere with retrieval functions. Information seeking at the level of scientific research is accumulative, each new finding supporting or questioning theories and principles. Information seeking at the operational level often applies results quickly and directly, archiving or discarding information as soon as it is applied, rather than making it part of the corporate memory. Because humans cannot selectively erase their memories, much of the information processed as part of information seeking is remembered regardless of whether we think it will ever be used again. As we depend more on external augmentation of our memories, especially through electronic technology, we can and should make conscious decisions about whether and how to store information. We must consider information from a life-cycle perspective where destruction options are developed along with generation, acquisition and storage options. As we seek, evaluate, and acquire information, we must consider integratability with respect to our

existing private or corporate knowledge and reusability for future problems. Technology can surely help us be more selective about what information we store and thus use our mental and external resources better, but we must balance optimization against our abilities to spout trivia at parties or make disparate connections that spark intellectual breakthroughs.

Learning takes place in directed and incidental ways but information seeking as defined here is a directed (purposeful) activity. There are, however, two ways that information can be acquired incidentally. First, our physical survival depends on our senses constantly gathering information about the environment to alert us to dangers and possible gratifications. This kind of automatic search for information is important to survival but is beyond the scope of our definition of information seeking. Second, as we purposefully seek information, we encounter many prospective units of information that are filtered and compared. Much of this irrelevant information is remembered automatically in spite of our efforts to ignore it and we are concerned with ways to minimize and label such information. The focus of this book is on purposeful information seeking, a process driven by an information problem. The information problem can be a mild curiosity, the desire to occupy 30 minutes, an ongoing passion about a hobby, or a desperate quest with life-critical consequences, but it must initiate conscious activity to move toward a goal.

As with learning or problem solving, we develop strategies to guide our progress. We use a variety of gross strategies in information seeking, including: consulting our own long-term memory; asking friends, colleagues or experts; consulting personal collections of books, periodicals and files; conducting empirical investigations; and applying formal systems. Formal systems include: libraries, research firms, government agencies, electronic networks, and the growing collection of information services that make up the information industry. The main focus of this book is on information seeking using formal systems, although as we shall see in later chapters, electronic technologies are blurring the distinctions between personal, informal and formal information systems.

In addition to strategies defined by what sources are used for search, there are strategies for how to search. A fundamental distinction is made between analytical and browsing strategies (Marchionini & Shneiderman, 1988; Liebscher & Marchionini, 1988). *Analytical* strategies depend on careful planning, on recall of query terms, and iterative query reformulations and examinations of results. *Browsing* strategies are heuristic and opportunistic, and depend on recognition of relevant information. Analytic strategies are batch-oriented, half-duplex (turn taking) like human conversation, whereas, browsing strategies are more interactive, real-time exchanges and collaborations between the information seeker and the information system. Browsing strategies demand less cognitive load in advance and more steady attentional load throughout the information-seeking process. Analytical strategies can be applied by intermediaries for

the benefit of the person requesting information, whereas browsing strategies are conducted by the ultimate user of the information. In practice, people apply different mixes of analytical and browsing strategies but electronic environments have severely constrained what strategies information seekers can use. Although people have an inclination to browse, analytical strategies are more efficient in large document collections. Early computer systems required analytical strategies and some present-day systems require browsing under the guise of ease of use. Both these developments have influenced how we seek information and this book lays out distinctions between these strategies and argues for a new generation of designs that support both strategies.

Figure 1 illustrates the relationships among learning, information seeking, information retrieval, and analytical and browsing strategies. Information seeking is often a type of learning since the goal in both cases is to change knowledge. Information seeking differs from learning according to the degree of retention desired; learning demands retention, information seeking may use the information for an temporary task. Much of information seeking may require identifying and retrieving previously stored information. Thus, information retrieval is one type of information seeking except when it is conducted by a machine--machines cannot engage in information seeking but machines can engage in information retrieval. Information retrieval, when practiced by people may be done to support learning. Browsing is often a type of learning and is not usually driven by well-defined goals or proceed according to systematic plan as is information retrieval. Analytical search is sometimes a type of learning and is most closely associated with retrieval. As will be demonstrated when these strategies are characterized in later chapters, browsing and analytical search present some similarities and their overlap.

[Place Figure 1.1 about here]

Other strategies we develop include applying filters or templates to search and to minimize overload, broadening or narrowing the scope of search, and rationalizing results regardless of relevance. Electronic environments have begun to affect the strategies we use in information seeking and these effects are considered in the chapters ahead. For example, because planning and query articulation are difficult, electronic systems that depend on recognition and interaction have become popular. A user-centered perspective on information seeking is taken in this book and arguments for systems that amplify such strategies are made.

PREVIEW

In the chapters ahead, we will consider how electronic environments affect information seeking and how interface design can support information seeking. Chapter 2 presents the perspective on humans and technology that serves as a context for the book. A general framework for information seeking is developed in Chapter 3 and examined in

subsequent chapters with respect to the effects of electronic technology. The framework is analogous to a computer program in that it includes a set of interrelated factors (data structures) and a network of processes (procedures). Factors include: information seeker, an information need manifested as a search task, domain, setting, search systems (information sources and interfaces, including people, books, computer systems, etc.), and search outcomes (the products and traces of information seeking). These components are managed by the information seeker as information seeking progresses. Information-seeking processes are the functions or actions taken during information seeking and include: problem recognition and acceptance; problem definition and clarification; source selection; query articulation; query execution; result examination; information extraction; and reflection, iteration, and termination. Chapter 4 focuses on the human factors that drive information seeking and facilitate information-seeking strategies. The framework is used in Chapters 5 and 6 to illustrate how information seeking in electronic environments compares to information seeking in manual environments and to examine analytical and browsing strategies in detail. The main differences discussed are related to speed of access; the scope of access (amount of information available); provision of interactive assistance and help during information seeking; flexibility and choice in representation of information; availability of powerful retrieval techniques such as string search and relevance feedback; physical constraints of using electronic equipment; changes in expectations; and overall high levels of interactivity that invite browsing and heuristic strategies (Marchionini, 1987). Analytical strategies developed for online searching are considered in Chapter 5 and browsing strategies are illustrated in Chapter 6 with interfaces from end-user-oriented systems. Chapter 7 considers directions for interface design that supports both types of strategies. Chapter 8 summarizes the effects on information seekers including: physical changes such as decreased movement from place to place and more focused hand, eye and muscle activity; cognitive effects such as cognitive amplification and augmentation; emotional effects due to increased interactions with systems replacing interactions with other people; and social/economic effects such as the evolution of personal information infrastructures and costs of systems, training, and information itself. It also considers the constraints on continued evolution of information seeking. The final chapter asks the reader to consider how information seeking should continue to evolve.

1. Shoshana Zuboff has used the more action-oriented term "informated society" in her book In the Age of the Smart Machine. Although this phrase is more appropriate to the highly interactive environments discussed in this book, the more popular term "information society" will be used. Drucker (1968) uses the term “knowledge society” and Bell (1973) describes the knowledge-driven, technological society as “post-industrial.”

2. Data from the 1990 U.S. census indicates that 46% of American children use computers in school or at home and that 37% of American adults use computers at work (Communications of the ACM, June 1991). The Statistical Abstracts of the United States, 1993 reports that 99.4% of all high schools have computers (mean of 45 per school) and 98.5% of elementary schools have computers (mean of 19 per school).

3. Examination of today's remote controls for home entertainment systems suggest support for the wag's observation that the tools we use to control our tools are more complex than those we use to build them.