# Information Seeking in Electronic Environments

Gary Marchionini
*University of Maryland*

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge CB2 1RP, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK http: //www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA http: //www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia



First published 1995
Reprinted 1996
First paperback edition 1997
Reprinted 1998

Printed in the United States of America

Typeset in Times

*A catalogue record for this book is available from the British Library*

*Library of Congress Cataloguing-in-Publication Data is available*

ISBN 0-521-44372-5 hardback
ISBN 0-521-58674-7 paperback

# Contents

# Figures and tables

## Figures



## Tables

# 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, largely because of developments in technology and the importance of information to human and technical development. We live in an information society[1] in which 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 acquire new knowledge to do their jobs. But new knowledge is no longer acquired only to prepare for a career but, rather, is now an essential part of "knowledge work." As Zuboff (1988, p. 395) 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. More than 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 increasingly more sources and larger volumes are 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 5500 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 million titles (*Informing the Nation: Federal Information Dissemination in an Electronic Age*). Not only are the volumes astounding, but they also are growing rapidly. For example, 90 billion pieces of first-class mail were handled by the U.S. Postal Service in 1991 versus 60 billion 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 are growing in size and complexity. The growth of directories and indexes is reflected in the 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 the incessant information assaults on our senses. To cope with these large amounts of information in our lives, we have developed 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 is printed or painted information. Much more information is becoming available in electronic form. The number of publicly available databases grew from 400 in 1980 to 8400 in 1993; the number of database producers grew from 221 to 3260; and the number of online services rose from 59 to 825 during this same period (*Gale Directory of Databases,* 1993). In addition, more than 3500 CD-ROM titles were available in 1993 (*CD-ROMs in Print 1993: An International Guide*). Although much of this information is also available on paper or microfiche, there is a burgeoning amount of information available only in electronic form. The collection of more than 31,000 computer networks (*Communications of the ACM,* August 1994), known as the *Internet,* reaches 20 million users worldwide and is growing exponentially. For

example, 7.5 billion packets were sent over the Internet in the month of May 1991, and this monthly volume had grown to 26 billion packets by February of 1993 (*Merit/NSFNET Link Letter* 6(1), p. 12) and to 56.2 billion packets by March 1994 (Merit Network Inc., personal communication, May 24, 1994).

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 mouse clicks. 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 use the computational power of computers systematically to 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), but 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 in many ways; for example, the set of bits 1001101 can be expressed on a display as an uppercase "M" or as a set of black and white pixels in a larger image, or as part of a note value for a compact disk 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 in which 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 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 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 created, 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 spurred advances in human–computer interface research. Although much 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 mean that for the foreseeable future, significant material and intellectual resources must be devoted to acquiring, learning to use, applying, and maintaining electronic tools. At the very least, it is obvious that more and more of our time and financial resources must be spent using computers, and we will become even more dependent on them in the future.

An important aspect of these effects is that more of our professional and personal lives will be 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 38 million users by 1995 (Reinhardt, 1993). Individual people 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, the automatic teller machines, phone menu systems, and information kiosks at shopping malls or museums.

As more of these systems are connected, 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 saves time and resources, and the computational power facilitates the execution of multiple tasks concurrently. The implications of mediated communication, high rates of exchange, and 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. In addition, the information-seeking process is more and more 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 depends 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 need 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* is used to refer to several different concepts in this book. Buckland (1991) 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 distinguished 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 can change a person's knowledge. This sense, according to 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 is used in this book 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 people's minds. To seek information, people seek to change the state of their knowledge and also physical representations (e.g., ink on paper, sound waves, electronically charged phosphorus) that represent abstractions (e.g., words, numbers, images, concepts, melodies) that can cause this change.

Because there are many manifestations of information-as-object, there are many terms that can be used to describe those objects. Terms such as bit, data, record, text fragment, graphic, document, utterance, database, book, and library all are used to label particular information units. Although these terms are typically associated with different media or information systems, the terms *document* and *information object* are used in this book 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 that we will call *information seeking,* a process in which humans purposefully engage in order to change their state of knowledge. The term *search* is used to mean the behavioral manifestation of humans engaged in information seeking and also 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, those people who "knew" it organized it for later "knowing" by themselves or someone else. Seeking connotes the process of acquiring knowledge; it is more problem oriented as 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 answering questions 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 more 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 help us generate, manipulate, and represent 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, the 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 help manage the generation, storage, and flow of information. Likewise, an enormous amount of information is necessary to select from entertainment options and to make good consumer decisions. 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 in order 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 stored so that it can be recalled and used at a later time, although information acquired as a result of information seeking may be useful for a specific task and then discarded. Inter-

mediate 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, with each new finding supporting or questioning theories and principles. Information seeking at the operational level often uses the results quickly and directly, archiving or discarding the 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 we will ever use it again. As we depend more on the external augmentation of our memories, especially through electronic technology, we should decide whether and how to store information. We must consider information from a life-cycle perspective in which 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 purposely seek information, we encounter many prospective units of information that we filter and compare. We remember much of this irrelevant information automatically in spite of our efforts to ignore it, and so we should be concerned with ways to minimize and label such information. The focus of this book is on intentional 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 a 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 among personal, informal, and formal information systems.

In addition to strategies defined by what sources are used for search, there are strategies for how one should search. A fundamental distinction is made between analytical and browsing strategies (Liebscher & Marchionini, 1988; Marchionini & Shneiderman, 1988). *Analytical* strategies depend on careful planning, the recall of query terms, and iterative query reformulations and examinations of results. *Browsing* strategies are heuristic and opportunistic and depend on recognizing relevant information. Analytic strategies are batch oriented and 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 a smaller cognitive load in advance and a steadier 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 limited 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 distinguishes between these strategies and argues for a new generation of designs that support both strategies.

Figure 1.1 illustrates the relationships among learning, information seeking, information retrieval, and analytical and browsing strategies. Information seeking is often a type of learning, because the goal in both cases is to change knowledge. Information seeking differs from learning according to the degree of retention desired; learning demands retention and information seeking may use the information for a 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 seek information but they can retrieve information. Information may be retrieved by people in order to support learning. Browsing is often a type of learning and is not usually driven by well-defined goals or does not usually proceed according to a systematic plan as information retrieval does. 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 do have some similarities.

Other strategies that we explore are 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

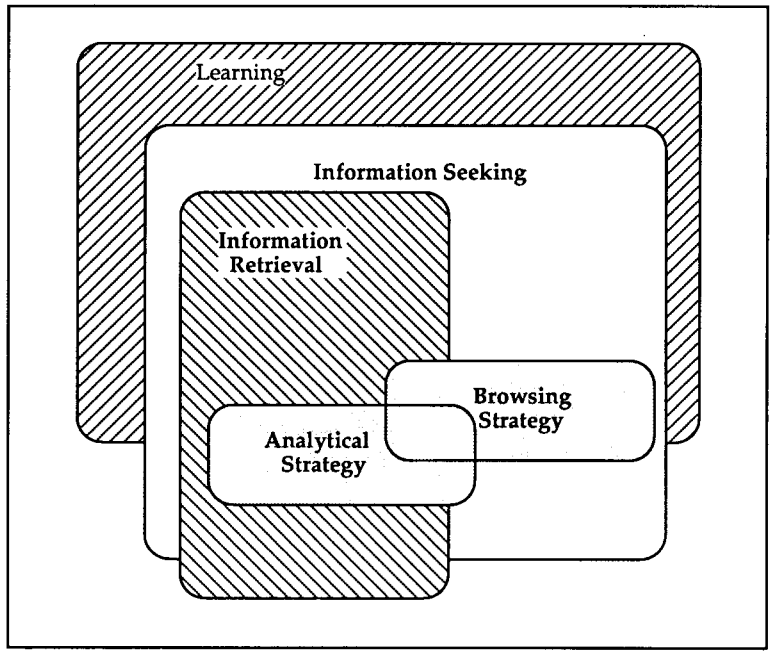


Figure 1.1. Relationships among key processes.

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. We take a user-centered perspective on information seeking in this book and argue for systems that amplify such strategies.

## Preview

In the following chapters, we 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 the context for this book. A general framework for information seeking is constructed in chapter 3 and is examined in subsequent chapters with respect to the effects of electronic technology. This framework is analogous to a computer program in that it contains a set of interrelated factors (data structures) and a network of processes (procedures). These factors include an information seeker, an information need manifested as a search task, a domain, a setting, search systems (information sources and interfaces, including people, books, computer systems), and search outcomes (the products and traces of information seeking). These components are managed by the information seeker as the information seeking progresses. Information-

seeking processes are the 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 with information seeking in manual environments and to examine analytical and browsing strategies in detail. The main differences that we discuss 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 representing 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 for online searching are considered in chapter 5, and browsing strategies are demonstrated in chapter 6 with interfaces from end user–oriented systems. Chapter 7 looks at the 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 and economic effects such as the evolution of personal information infrastructures and the costs of systems, training, and the information itself. It also discusses the constraints on the continued evolution of information seeking. The final chapter asks the reader to think about how information seeking should continue to evolve.