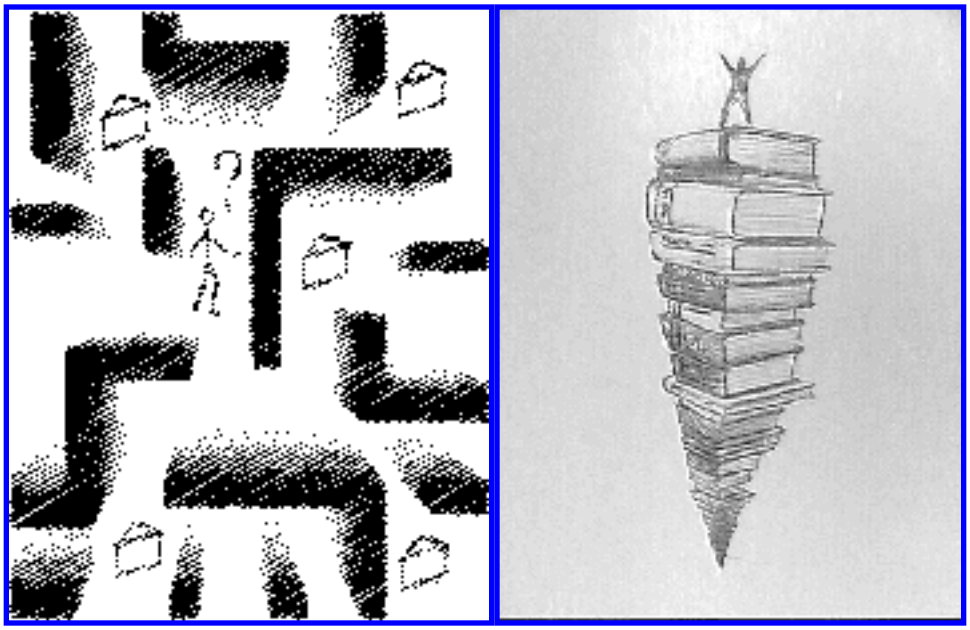
Crossroads
The ACM Student Magazine

We Have the Information You Want, But Getting It Will Cost You: Being Held Hostage by Information Overload

Mark R. Nelson

The Information Age has arrived and with it comes a daily assault of increased information. Society is being held hostage by a battery of information which threatens to exceed our ability to manage it. Information overload costs businesses and individuals valuable time, effort and additional resources...and the cost is rising.



Wurman [20] writes, ``a weekday edition of *The New York Times* contains more information than the average person was likely to come across in a lifetime in seventeenth-century England.'' In today's society, the success and survival of many companies and individuals hinges upon their ability to ``locate, analyze, and use information skillfully and appropriately [10].'' Our proficiency at generating information has exceeded our abilities to find, review and understand it. The problems associated with the explosion of information are exemplified on the ``Internet-turned-Information-Superhighway.'' The volume of information on the Internet has exceeded the ability of

most people to find the information they need. The tools to support resource identification and use have not increased in effectiveness as rapidly as the quantity of available information has increased.

Getting the Right Information at the Right Time

In settings like the Internet, there are a wide range of information sources located in many different places. In order to more effectively locate the information which is most useful, there are a number of tools and centralized resources that might be useful. To reduce confusion when searching for information, a person should know what tools are available and how to use them. Information on the Internet is posted in a number of different formats located at thousands of sites. Information is stored in and across sites and can be referenced geographically, topically, by keyword or by random chance and exploration.

Information can be found in bulletin boards, Usenet news groups or within ``gopher-space''. One problem with many Internet resources is that most are hard to find. There are many resources available on the Internet, but finding them is not easy using the tools at hand. Schwartz [18] talks about the importance of resource discovery. He emphasizes that ``if users cannot discover resources, they get only a fraction of the potential benefits of a network.''

One reason that resource discovery is difficult on the Internet is that many resources are not ``advertised'' well or are difficult to access. Additionally, Schwartz [18] points out that most resource discovery systems do not distinguish between discovering some class of resource (such as job postings in general) and locating an appropriate instance of a resource (such as a listing containing specific job openings on a particular Internet host). It is quite possible that some information might be found in multiple locations. A user would probably not want to peruse the same posting twice, or the user may need to determine which copy of the information is the ``accurate'' version.

Defining Information Overload

What is information overload? To answer the question, there must be a clear definition of information itself. This article defines ``information'' as ``data'' or ``facts''. As will be seen in later sections, this broad definition will simplify some of the problems which comprise the information overload dilemma.

Information overload is the inability to extract needed knowledge from an immense quantity of information for one of many reasons. Wurman [20] explains that information overload can occur when a person:

- does not understand available information.
- feels overwhelmed by the amount of information to be understood.
- does not know if certain information exists.
- does not know where to find information.
- knows where to find information, but does not have the key to access it.

Although many problems surface as part of the information overload predicament, this article will focus only on a subset of issues. This subset includes problems with the accessibility and volume of information, usability of information sources, and what is meant by information anxiety. Each of these concerns is involved with the issues of information quality and accuracy.

The ``Big'' Information Problem: Volume

Perhaps the first thing which comes to mind when thinking about information overload is the quantity of information that must be examined. Often, we sort through large quantities of information just to locate the small piece that we need. Naisbitt [14] writes, ``Inundated with technical data, some scientists claim it takes less time to do an experiment than to find out whether or not it has been done before.''

Naisbitt's comment may seem to be a little dramatic, but perhaps it is more correct than we care to admit. Murray [13] estimates the following:

> In every 24-hour period approximately 20,000,000 words of technical information are being recorded. A reader capable of reading 1,000 words per minute would require 1.5 months, reading 8 hours every day, to get through 1 day's technical output, and at the end of that period, he would have fallen 5.5 years behind in his reading!

Consider that more new information has been produced within the last three decades, than in the last ***five millennia*** . Over 9,000 periodicals are published in the United States each year, and almost 1,000 books are published daily around the world [11]. The November 13, 1987 issue of *The New York Times* [20] numbered 1,612 pages, containing about 2,030,000 lines and over ***twelve million*** words. Is it any wonder that the term ``*information explosion''* has become so commonplace?

One result of having all of this information available, is an increased difficulty in finding the particular information for which we are searching. We have to determine which information is useful, which is not, and where to look next when necessary. In one article, Fine quotes a piece by Newman which explains the problem well. Newman [5] says:

> technology is ***volume*** -- a greater number of data, more materials, more items, more detail. The result is that sometimes we are provided with both useful and useless information, and we must learn quickly to sort and choose.

Fine and Newman [15] distinguish between information and ``real need knowledge.'' There is a difference between what information is available to us and what information we need or can use. A recent posting from a Usenet News Group illustrates the point in another way:

> There is something about [the Internet] that generates volumes of talk without necessarily increasing the quality. There are just too few nuggets to be mined from this river of information. And I realize that one person's nugget is another person's worthless pebble. But the bigger the network becomes, the more pebbles each individual must sift through. Even at cocktail parties I can be selective about the conversations I drop into. Here they all drop in on me. [Quoted from a posting to the LM_NET discussion group by sgrant@eis.calstate.edu]

If all information was of equal value to everyone, controlling the volume in a more meaningful way might be easier. Unfortunately, this is not the case and so a person must search for personally meaningful information. To reduce the amount of time used to search for ``real need knowledge'' we must find a way to successfully overcome this aspect of the information overload riddle.

Verification of Information Accuracy

The volume of information on the Internet creates more problems than just trying to search an immense collection of data for a small and specific set of knowledge. Large volumes of data are fraught with inconsistencies, errors and useless data. When we try to retrieve or search for information, we often get conflicting information or information

which we do not want. Therefore, validating information is another important aspect of information overload.

Individuals searching for information want to maximize the number of items found that are highly relevant to their work. Curtis and Rosenberg [3] contend that a search request consists of two components which determine the accuracy of search results. The first is the point of view of the person seeking information and the second deals with system functionality.

Users must take on the responsibility of designing and structuring the parameters of their searches to match their own points of view. Depending on a user's ability to select search parameters, one hopes to eliminate a large portion of *false drops*, while not excluding relevant documents. False drops are those documents which match given search criteria, but are irrelevant to the user's needs. According to Curtis and Rosenberg [3], an individual must be willing to sacrifice the possibility of ``instantaneous response'' for greater accuracy of retrieval.

Accuracy of information content is an important factor in the validation process. Schwartz [18] mentions two causes of inaccurate information: outdating and inconsistent copies between distributions. Information often must be timely if it is to be useful. In many cases ``old'' or outdated information has as little value as no information at all, or worse, may have negative value.

Similarly, if one finds two copies of a file, but they do not match, how does one determine which version is the accurate one? Whose responsibility should it be to verify the authenticity of data, or what happens if a person finds only one copy of a file, but it is an incorrect or inaccurate version? Often, conflicting copies of information proliferate as duplicate research is performed, recorded and posted [8]. On the Internet, the person who retrieves a piece of information is most often responsible for determining its accuracy. The need to validate information simply increases the bulk of information to be processed by the user.

Information Literacy and Application Usability

There appear to be two major factors affecting the ability of people to access information effectively: information literacy and application usability. Information literacy is ``the ability to effectively access and evaluate information for a given need [2].'' Application usability refers to the interactive environment which a software

application or system provides to a user searching for information.

Information literacy is the ``people'' aspect of information access. Horton [9] described the purpose of information literacy as:

> ...raising the levels of awareness of the knowledge explosion and involving understanding as to how computers can help identify, access, and obtain data and documents needed for problem solving and decision making.

Being able to understand what is required to find information is an important element in the process of overcoming information overload. Some skills which contribute to information literacy are problem solving, decision making, critical thinking, information gathering and interpretation [2]. These skills are needed in addition to a basic competence and familiarity with computers. The application usability side of information access requires that computer-based information systems be designed for ease of use. Important interface concerns include selection methods (command languages versus menus), and representation methods (screen layout and graphic/text combinations) [12].

The most popular models [12,1] used in measuring the effectiveness of an information retrieval system contrast ``*recalling* with *recognizing*'' with ``*recall* with *precision*.'' The basic premises of these models involve the following facts respectively:

1. humans have a working memory limited to 5-7 ``chunks'' of information;
2. humans must have their attention refreshed frequently;
3. and RECALLING information requires more cognitive effort than RECOGNIZING information [12].

RECALL is the proportion of all relevant documents that are actually retrieved. PRECISION is the proportion of a retrieved set of documents that is actually relevant [1].

The first model suggests the system should accommodate human limitations, maximizing the need to recognize useful information as opposed to recalling it. This would indicate that systems which use menus are more effective and user-friendly than those which require knowledge of a particular command language [12].

The second model considers how accurate a system is at retrieving relevant

documents. A system is not very usable if one must exert significant effort to retrieve relevant documents or discard irrelevant ones. Accurate document selection is an important factor in determining the usability of a system [1].

Finally, the combination of usability and information literacy yields enhanced productivity. Certainly, workers are more productive if they use a product which is ``well liked, easy to learn and contains the right functions. [7]'' A vital component in the pursuit to free ourselves from information overload will be the development of applications that are more user-friendly and precise.

Current Trends

The trends in information retrieval practices are probably as numerous as the types of information they access. However, there are some particular trends in the battle against information overload which deserve a few remarks here. Some key trends involve hypertext systems, gophers and natural language processing systems. These selected trends represent only a small portion of the myriad resource discovery services in place or under development. Hypertext uses the principles of recalling and recognizing discussed earlier [12]. Hypertext is a software framework which represents textual information in a non-linear fashion, allowing user-directed navigation [17]. The information is stored in *nodes* (concepts) and connected by *links* (associations). A node contains information or text about a certain topic or concept. A link connects two nodes which are in some way related. Users follow the association links between concepts to find the information for which they are looking. For example, an earlier section discussed the principles of ''recalling'' and ''recognizing''. In a hypertext system, these terms might be linked to text explaining more about what the principles mean.

Like hypertext systems, gophers allow users to navigate links in order to find information. In ``gopherspace'' the links generally exist among menu items and resources rather than in concept pairings. Gophers organize information into a hierarchy of menus. The leaf nodes in the hierarchy are either resources or documents. Each of the intermediary nodes are directories or indexes. Gophers are very powerful as they allow users to retrieve information from many of the different resources and repositories available on large scale networks such as the Internet [15].

Perhaps the biggest downfall of gopherspace on the Internet is that it is still hard to find many sources of information. If a person knows the location of a particular resource it is easy to find. However, finding general information, such as job postings

can be quite difficult. An exploration of gopherspace will yield a listing of over 900 gopher sites spread around the world. Since many sites offer specific services, a person might have to check them all to see if they have a particular item, resource or document of interest.

Most traditional approaches to information retrieval involve searching for the occurrence of a particular set of indexed terms in order to find relevant information. These systems often return unwanted information since they have no real understanding of the information itself. In other words, the information means nothing to the application which accesses it [16].

Natural language processing systems use artificial intelligence methodologies to search for relevant information. These systems generally find information based on the relationship between the content of the available data and the user's interests. Natural language processing systems use sophisticated techniques to search and prune their information set and return relevant information to the user [16].

Natural language processing systems hold much promise for assisting with the information overload problem. Consider a scenario where an employer is searching for employees from a pool of applicants whose resumes are in an electronic format. The employer could ask the system questions like: ``Are there any applicants with a graduate-level degree in Computer Science or related field who have three or more years of work experience?'' A job seeker might ask questions like: ``What computer jobs are available in Georgia?'' or ``What management positions are available at ABC Company?'' These systems have the potential of being easy to learn and user-friendly while still being very powerful.

Conclusions

Information anxiety is the overwhelming feeling one gets from having too much information or being unable to find or interpret data. Wurman [20] writes:

> Information anxiety is produced by the ever-widening gap between what we understand and what we think we should understand. It is the black hole between data and knowledge, and it happens when information does not tell us what we want or need to know.

Information anxiety is the primary defining characteristic or result of the information

overload problem. If a person did not have any problems finding the correct information, or if the information came in just the right quantity, then information overload would not exist. Information anxiety results from our inability to access and extract meaning from the wide accumulation of information available to us. Williams notes, information itself has no value, it is the communication and sharing of information and its meaning which gives it value [19].

There needs to be a better way to manage information without being held hostage by it. If we improve how we retrieve information, then our information anxiety will be reduced. Denning [4] described it as:

> The visibility of personal computers, individual workstations, and local area networks has focused most of the attention on **generating** information - the process of producing documents and disseminating them. It is now time to focus more attention on **receiving** information - the process of controlling and filtering information that reaches the persons who must use it.

In short, there are many places to look for information, but some resources are more valuable or easier to use than others. As the issues of information volume and accuracy become greater, the need for new and better tools to handle these problems will also increase. As described by Wurman [20], ``The greatest crisis facing modern civilization is going to be how to transform information into structured knowledge.'' Society faces an over-abundance of data that needs to be evaluated and acted upon. Whether searching online data repositories or gopherspace, a person must find a faster and more effective means of retrieving information. Individuals require easier access to the many resources which are available, but are hard to find. The result of better information retrieval systems will yield significant benefits to everyone.

References

Belkin, N. and Croft, W. B. Filtering and information retrieval: two sides of the same coin? *Communications of the ACM*. 35, 2, (December 1992).

Breivik, P. Putting libraries back in the information society. *American Libraries.* 16, 11, (November 1985).

Curtice, R. and Rosenberg, V.. *Optimizing Retrieval Results with Man-Machine*

Interaction. Center for the Information Sciences, Lehigh University, Bethlehem, PA. February 1965.

Denning, Peter. ACM president's letter. *Communications of the ACM*. 25, 3, (March 1992).

Fine, Sara. *Information Technology: Critical Choices for Library Decision-Makers.* ed.

Kent, A. and Galvin, T. New York: Marcel Dekker, Inc. 1982.

Gier, Sue. Relief from information overload. *The Computer Connection*. James Madison University, VA. November, 1990.

Gould, J., Boies, S. and Lewis, C.. Making usable, useful, productivity - enhancing computer applications. *Communications of the ACM*. 34, 1, (January 1991).

Hammer, Donald. National information issues and trends. *Annual Review of Information Science and Technology*. (American Documentation Institute. Annual Review Series, vol. 2). Interscience, New York. 1967.

Horton, F. Information literacy vs. computer literacy. *Bulletin of the American Society for Information Science*. 9, 4, (April 1983).

Hubbard, Susan. On Kuhlthau, C.C.: *Information Skills for an Information Society: A Review of Research*. Syracuse, NY: ERIC Clearinghouse on Information Resources. December, 1987. ERIC number ED 327 216.

Large, Peter. ***The Micro Revolution Revisited.*** New Jersey: Rowan and Allenheld Company. 1984.

Marchionini, Gary. *Psychological Dimensions of User-Computer Interfaces*. Syracuse, NY: ERIC Clearinghouse on Information Resources. ERIC no. ED 337 203. October, 1991.

Murray, Hubert, Jr. *Methods for Satisfying the Needs of the Scientist and the Engineer for Scientific and Technical Communication* in a press release, March 8, 1966. Washington, D.C..

Naisbitt, John. *Megatrends*. New York: Warner Books. 1982. (This book is out of print, but you might be interested in a newer version: **Megatrends 2000** by the same author.)

Obraczka, K., Danzig, P. and Li, S.. Internet Resource Discovery Services. *IEEE Computer*. September, 1993.

Ram, Ashwin. Natural Language Understanding for Information-Filtering Systems. *Communications of the ACM*. 35, 12. (December 1992).

Schroeder, Eileen. Interactive Multimedia Computer Systems. Syracuse, NY: ERIC Clearinghouse on Information Resources. ERIC no. ED 340 388. November, 1991.

Schwartz, Michael. Internet Resource Discovery at the University of Colorado. *IEEE Computer*. September, 1993.

Williams, Martha. Online Retrieval -- Today and Tomorrow. *Online Information Systems*. Oxford, England: Learned Information Ltd. 1984.

Wurman, Richard Saul. *Information Anxiety*. New York: Doubleday. 1989.