VICTOR ROSENBERG

President
Personal Bibliographic Software, Inc.
Associate Professor
University of Michigan-Ann Arbor

Desktop Research and Software Connectivity

ABSTRACT

"Desktop research" encompasses the various tools that a scholar requires in the course of his or her work. The "scholar's workstation" of the future will involve several software packages from a number of developers to accomplish the tasks required in doing research and creating publications. To function effectively, the programs must be able to interact with each other and communicate data. A common user interface will ease the learning of each new addition to the software repertoire. A model workstation is discussed that allows searching of bibliographic databases or library catalogs, the assembly of bibliographies, the ordering and acquisition of documents, and the preparation of manuscripts. (The workstations to support the concept of desktop research were provided under an Apple Library of Tomorrow grant from Apple Computer, Inc.)

SOFTWARE CONNECTIVITY

A great deal of attention has been paid in recent years to the connections between computers of various sorts. This connectivity includes the use of microcomputers to access mainframe or supercomputers as well as the connection of a number of microcomputers into a local area network or workgroup set of computers. The advantages of connecting computers and establishing communication between them is obvious. But the connection of two computers whose users are working

with incompatible software is like two speakers in a telephone conversation speaking different languages. While there is a physical communication, the substance is entirely lacking.

Software connectivity is the ability of various software packages to work together in such a way as to make the whole greater than the sum of the parts. Although software vendors, as well as vendors of information, would like to believe that a user will use only their products, it is becoming increasingly obvious that most computer users will be working with a variety of software packages, often to accomplish a single task, and they will require access to many different sources of data. One answer that the software industry tried was integrated software, but this never succeeded to any great extent. No one software company could achieve excellence in all the necessary software products, and different combinations of products are necessary for many applications. Now, more and more vendors are designing their software with the necessary capabilities to work with other vendors' software. The result is the ability to import and export data readily between applications and thus the ability to process data through several products in sequence. There are even some more or less standard ways to handle certain graphics and textual data. For example, any spreadsheet package can read or write Lotus 1-2-3 files. This means that any spreadsheet program can communicate with any other spreadsheet package using the Lotus "standard" as an intermediate format.

Even more important is the realization that a standard user interface for a large number of programs will reduce the learning necessary to add software to an individual's repertoire and make the transition from one software product to another easier. As encouraging as these developments are, there are also counter-trends. For example, Apple Computer, long the leader and champion of the standard interface, has introduced HyperCard, a computer programming language that makes the interface the subject of the software author's whim. Although it is possible to adhere to the Apple standard using HyperCard, most HyperCard authors cannot resist the temptation to create an innovative interface and perhaps even set yet another new standard. Similarly, in the MS DOS and UNIX world, there still has not been any substantial agreement on what a standard interface should look like. A number of companies are attempting to create the standard interface.

This fierce battle over the interface is an indication of how large the stakes for the winner are. Clearly, computer and software manufacturers realize the gains to the developer of the standard interface are enormous, and thus each manufacturer is struggling to establish its own interface as an industry standard. Several companies are even getting together to develop a standard while others are filing lawsuits against each other over ownership of the interface. The stakes have

to be high to spawn that much cooperating as well as fighting among competing organizations. (It has been said that the nice thing about standards is that there are so many to choose from!)

Even without the utopia of a common interface and clear data format standards, it is still possible to connect several software packages together and even transfer data between different computers with different operating systems. It is possible to exchange word processing documents between several products on a single computer and to exchange documents between major word processors on different machines. For example, with Microsoft Word or WordPerfect, it is possible to transfer a document between an IBM PC and a Macintosh. Pro-Cite can open a database on an IBM machine from the Macintosh and vice versa. Databases can be transferred back and forth. Soon, UNIX systems will have the same capability with Pro-Cite. This interchange of data between machines and software packages makes it possible for people with different machines to work together and makes the purchasing choice of which personal computer to buy a bit less harrowing.

The "Scholar's Workstation"

An example where software connectivity can play a major role is in the work of a campus researcher. In the university of the future, a student or faculty member embarks on a research project at an advanced workstation. One example might be a medical student working on a paper dealing with a new drug treatment for AIDS. He or she begins by searching three sources for bibliographic information.

The three sources to be searched are MEDLINE on a CD-ROM player next to the personal computer, Biosis and Chemical Abstracts databases on the Dialog Information Services, and the local university library's online catalog. From these three sources, the student will assemble a collection of references on the topic. Most citations will include abstracts. The CD-ROM is searched using one of the many providers of MEDLINE on CD-ROM. The Dialog databases and the online catalog are searched using Personal Bibliographic Software's Pro-Search. The combined records are downloaded and converted to a database in the workstation using Pro-Cite and Biblio-Links. Duplicate records are eliminated, and Pro-Cite is used to produce a bibliography for the paper in the Council of Biology Editors format required by the journal to which the paper will be ultimately submitted.

From the Pro-Cite database, the student can select the documents he or she wants to examine. This selected set is then sent by modem to a workstation in the library. The resulting documents are physically taken from the shelves and the relevant pages scanned into the library's

workstation. These images can then be sent via fax modem to the student's PC. The student will store the document images on the hard disk of the workstation. Using optical character recognition (OCR) software, the articles will be converted into ASCII characters. The student will then use a word-processor to begin work on the paper. Quotes from the scanned documents will be inserted into the document and citations of the form "(Smith, 1980)" will be appended to the quotations used. Charts and illustrations will also be cut and pasted into the paper with proper attribution. When the paper is nearing completion, a bibliography will be generated automatically and appended to the paper. When complete, the paper will be sent via modem to the student's professor. The paper will go directly to the professor's computer where it can be examined. It will also be printed using a laser printer.

This scenario suggests how the student or faculty member of the future will do library research and write the resulting paper. If laboratory work is a part of the research, the results of the experiment can be manipulated by computer and ultimately integrated into the paper. Since the intellectual property implications of this process are not yet fully understood, only public domain documents, or documents where appropriate royalty has been paid, can be used.

CONCLUSION

Implications for the Library, the Publisher, and the User

What are some of the implications of such a scenario? First, the student or professor does not have to set foot into the library to get relevant citations and does not even have to go to the library to obtain the needed documents. The current model involves the removal of the paper document from the shelf and the electronic scanning of the materials, a labor-intensive manual process. Ultimately, the library will subscribe to a journal subscription that does not exist on paper, but rather on a master file server at the publisher. When this happens, the documents need not be scanned to be sent to the student's workstation, because they already exist in that form on the server. The library will then function as a "switch," routing the student's document request to the appropriate server where the library has a subscription. In this scenario, students search online databases themselves and own the latest CD-ROM databases needed for their research. In fact, the same CD-ROM player used for bibliographic research can double as a music CD player that plugs into a stereo set!

What, then, is the function of the library, other than to function as a museum for books? The logical answer is that the library will still have to provide the reference function of directing researchers to sources, and this function will become increasingly sophisticated. But the most important function of the library will be education and training. The technology will require effective training and the library will be the logical place for this training function. In addition, the library will increasingly take over many of the functions now in the hands of the computing centers, i.e., the maintaining of the communication and computer equipment and software. The library is already the largest database in most universities and research institutions, and that will continue even with the new technologies.

The new technologies will have profound effects on publishers as well. They will no longer have to cut down trees to produce paper copies of books and journals. Mailing costs will be reduced, and virtually all materials handling problems will disappear. Problems of preservation will become moot as well, since digital information is infinitely replicable without image degradation. Of course, the problems of "information overload" will be worse than ever. The amount of information accessible to any scholar will be many times what it is now, and he or she will still have to sort it out and sift out all unwanted materials. The technology for the management of information in its physical form is vastly outstripping current ability to retrieve important information from the vast quantity of material in the universe of information.

Since the technology described above allows text to be converted from paper to electronic form and then transmitted, there are severe problems regarding the question of ownership and control of the information. Non-copyrighted material in the public domain is no problem, but proprietary information cannot be used without the permission of the owner of the copyright. Getting the permissions may prove to be more difficult than getting the documents.

REFERENCES

Glossbrenner, A. (1989). The universe of on-line information. *Macintosh Business Review* (January), 48-55.

Lucier, R. E.; Matheson, N. W.; Butter, K. A.; & Reynolds, R. E. (1988). The knowledge workstation: An electronic environment for knowledge management. *Bulletin of the Medical Library Association, 76*(3), 248-255.

McInnis, R. G., & Symes, D. (1988). David Riesman and the concept of bibliographic citation. *College & Research Libraries News, 49*(5), 387-399.

Moran, B. B.; Surprenant, T. T.; & Taylor, M. E. (1987). The electronic campus: The impact of the Scholar's Workstation Project on the libraries at Brown. *College and Research Libraries, 48*(1), 5-16.

Ofman, P. (1987). Personal bibliographic software: Rounding up those references. *Profiles* (March), 33-36.

Young, P. H. (1989). Library research in the future: A prognostication. One librarian's speculative bibliographic scenario. *College & Research Libraries News, 50*(1), 7-10.

Pitkin, G. M. (1988). CARL's latest project: Access to articles through the online catalog. *American Libraries, 19*(9), 769-770.

Weiskel, T. C. (1988). The electronic library: Changing the character of research. *Change, 20* (November/December), 38-47.

Wieskel, T. C. (1988). University libraries, Integrated Scholarly Information Systems (ISIS), and the changing character of academic research. *Library Hi Tech, 6*(4), 7-27.