M. K. PARK, M. C. CAUGHMAN, AND H. J. HAMILTON

Education in the Use of Modern Information Retrieval Techniques*

M. K. PARK†, M. C. CAUGHMAN, and H. J. HAMILTON
Computer Center, University of Georgia, Athens, Ga. 30601

Received January 18, 1971

The Information Center at the University of Georgia (UGA) has taken a multipronged approach to educating its users in the use of computer-based information retrieval services. For campus users, both seminars in the academic departments and personal interviews and discussions are used. A series of classroom lectures, augmented by practical exercises, has been initiated in conjunction with the chemical literature class. Workshops, which provide in-depth training on profile coding techniques and analysis and use of search results, are offered several times each year. An on-line, interactive search system, used to augment all of the above methods, is available for use from a CRT console for representative subsets of various data base files.

Every scientific researcher and teacher is aware of the rapidly increasing volume of scientific information. It is evident in the increasing amount of time he has to spend browsing through the periodical literature to keep abreast of the current developments in his area of interest. Initiation of research projects, preparation of lectures, and preparation of papers, reviews, and books all require extensive library resources and hours of time. Many government agencies, private research institutions, and major industrial organizations have formed information service departments to assist internal staff in obtaining desired information. But the university, in most cases, has left the task of information pursuit to the individual. Traditional library resources are available, but university libraries are usually not staffed to assist users with their library research beyond the shelf location of documents. And the university, as the training ground for scientific researchers, has, for the most part, continued to teach the traditional methods of using the literature.

The fairly recent application of computer technology to the handling of bibliographic information has opened up a whole new range of library services, and the University of Georgia has undertaken a fairly extensive program to make these resources available to its faculty, research staff, and students. Information such as the titles, bibliographic citations, abstracts, and indexing terms which are recorded in machine-readable form for editing are available as a by-product for computer searches. Tape services are now available in chemistry, biology, medicine, nuclear science, engineering, physics, psychology, geology, technical report literature, and education, to mention a few. Many smaller, specialized data bases have also been created, but their availability is generally limited, and, consequently, they do not constitute the library resource that the large data bases do.

The advent of computer-based information services has not been a panacea for information retrieval problems. It causes a change in the information use habits of the users. Perhaps it is too obvious to say that university users must be educated to both the capabilities and limitations of computer retrieval in order to make effective use of the services. Several papers have reported various aspects of this training,[1-5] and two related papers were a part of this symposium.[6,7]

The impact of the new services on present information habits of university personnel cannot be ignored, either. Time formerly spent in paging through periodicals is now available for more thorough analysis of the literature identified through the computer search. In fact, more time may actually be spent using the literature when computer-based aids are available than was previously spent with more traditional literature methods. This is especially true in the university community where researchers have tended to depend heavily on the so-called "invisible college." With computers doing the searching, literature which would otherwise be missed is brought to the users' attention, and with it comes the nagging responsibility to do something about it. We are aware of several individual cases of rather drastic changes in information use habits among the university user group. Several faculty members who initially used the computer searches to augment their manual searches and journal perusal have completely reversed the process. Their journal reading is now only their professional browsing, and they rely entirely on the computer retrieval for their searches. Just how widespread this trend is, however, has not been studied in detail in our user environment.

The Information Center at the University of Georgia has taken a multipronged approach to education of its user group. It is a broad program designed to appeal to many different types of users at several depths. It includes seminars, personal interviews, workshops, classroom lectures, trial search runs, and use of an on-line retrieval system. The educational program embraces the full range of services and data bases available in the

* Presented before the Divisions of Chemical Education and Chemical Literature, Symposium on the University and Chemical Information, 160th Meeting, ACS, Chicago, Ill., September 1970.

Information Center. This includes fourteen bibliographic data bases—not only in chemistry and biochemistry, but also in many other disciplines as mentioned earlier. The Information Science group does attempt to reach people in all these disciplines, but the program is more concentrated in chemistry than in any other area.

SEMINARS

On the Athens campus and at the other 25 colleges and universities in the University System of Georgia, the principal method of introducing the search services has been through departmental seminars for the faculty, research staff, and graduate students. These talks are usually 1 to 1½ hour presentations made as part of the regular seminar series of the host group. In general, the searches are presented as a library service, although the Information Science Unit has no official link with the University Library. Emphasis is placed on the characteristics and contents of the data bases, which are available in the particular subject area of interest to the group. The subject scope is explained in some detail, especially where several data bases are available which overlap each other to varying extents. Even though many of the attendees have used the printed forms of the publications, they seldom have a good understanding of the coverage scope of these abstract or index journals. The techniques of coding search profiles are described only in a very superficial way, just enough to give a general idea of the processes required to convert questions into computer-readable form. Examples of search results are presented and explained so that the users will understand the results and know how to interpret them. All presentations are made with visual aids in the form of slides, a complete set of which is available for each data base. Professional society meetings, both local and national, have also proved to be a good forum for reaching large groups of potential users.

USER INTERVIEWS

The seminar attendees and other prospective users of the system are encouraged to set up interviews with the information scientist who specializes in his subject area—e.g., chemistry, education, agriculture, etc. These personal talks have proved to be one of the most productive methods of educating the users to the real capabilities and limitations of computer-based services. In a two-way discussion, it is possible to describe those features of the service that are of specific interest to the user. He helps with the construction of his search profile to the extent of suggesting terms that are widely used in the related literature, providing the names of authors who publish in the field, and determining the relative importance of the various terms selected for the profile. Most users take enough personal interest in the profile to feed back suggestions for modifications after they have reviewed the results for the first few searches.

The primary factor in the success of this approach seems to be the fact that both the user and the information scientist talk the same language. All the information scientists in the unit have degrees in the subject disciplines for which they code profiles. When the computer-based services were first installed in 1965, the Computer Center had no personnel on its staff for this purpose; users were expected to read the profile coding manual and construct their own profiles. This first approach was a total failure. The users neither understood nor had the time to study the data bases and the coding conventions, and the Computer Center had no personnel on its staff who could train the users. As a consequence, the profiles performed very poorly; users lost all interest in the service; and, in fact, searches were discontinued for a time. In 1968, a chemist and a biologist, as well as information specialists and computer programmers, were added to the staff. During the last 2 years the program has grown from an initial 11 profiles against one data base to over 2000 current-awareness profiles. Retrospective searches are also offered on the accumulated files for all data bases. The program is continually being expanded with the addition of new data bases and corresponding staff to code profiles, primarily due to a grant from the National Science Foundation to support the research and development portion of the program.

CLASSROOM EXERCISES

A more formal educational project has been carried out for the past 2 years in conjunction with the chemical literature course offered by the chemistry department.[8] As a part of this project, four classroom sessions were devoted to instruction in the use of the computer-based services. The first three sessions were held during successive class periods. The fourth session was held about 2 weeks after the third session.

During the first three periods, a general description of the chemical data bases available and of profile construction techniques for *CA-Condensates* was presented. At the end of the third class period, each student constructed a profile on an assigned topic. During the interim between the third and fourth class periods, the students' profiles, together with those of an information scientist on the same topics, were searched against one volume of *CA-Condensates*. Each student also did a manual search of the keyword indexes in the corresponding volume of *CA*.

The fourth class period was devoted to reviewing the results of the three searches—the manual search, the student's profile, and the instructor's profile. The necessity of profile revision was emphasized. Although the students did not actually revise and research their profiles, it was felt that the comparison of their profiles with those of an information scientist was adequate to show the results of profile revision.

In other departments where no literature courses are offered, the responsibility for teaching students to use the literature usually rests with the graduate research adviser. In some cases the faculty member runs one or more profiles for his research group as a whole and shares the search results within the group. There is a growing trend, however, for the graduate advisers to require their students to visit the Information Science Unit and perform a literature search using these facilities before starting any new research for a masters or doctoral thesis. It is difficult to measure how widespread this practice is, but the unit is seeing the effect in the growing number

of students who request searches. Also, professors and instructors in departments ranging from library science to landscape architecture are increasingly using the information science staff as visiting lecturers for their classes.

WORKSHOPS

For users who wish to learn the details of profile coding, the Information Center conducts workshops several times each year. The workshops are 2½ days in length, divided into 5 half-day sessions. Attendees are assigned to small working groups based on their subject area of interest. The first session is introductory in nature, covering the basic features of computer-based retrieval, the characteristics, content, and editorial policies of the various data bases, and the basic requirements for construction of a search profile. Subsequent sessions are devoted to search result analysis, profile revision, and refinement techniques such as term weighting. Three search runs are made against each data base during the workshop. The results are then returned to each attendee at the following session for review of the results in conjunction with corresponding printed copies of the data base. After each review, the profiles can be modified for subsequent search runs. Workshop sessions have been held for both the bibliographic text data bases and the chemical structure notation files.

Workshops are limited to 40 attendees in order to assure personal attention and help to each participant. Most of the attendees are from industry and from government agencies, usually technical librarians or technical information managers. However, the sessions have included bench scientists, librarians, computer programmers, and teachers. University personnel, both from the University of Georgia and other universities throughout the country, have attended. In some cases, especially for industrial organizations, the person attending will be responsible for coding questions on one or more data bases in his own organization. In others, the technical librarian will serve primarily in an advisory, educational, and liaison capacity between the user and the Information Center who runs the searches. The producers of the data bases, themselves, often send representatives, usually to give their personnel first-hand experience in how the data bases are being used for retrieval services.

ON-LINE SEARCH SYSTEM

Perhaps the single most useful method of educating large numbers of people quickly has been the use of an on-line, interactive search system for several of the text data bases. With this system, search questions can be entered directly at a typewriter or cathode ray tube terminal, searched against the computer files, and the answers returned to the terminal within a few seconds. Four data files are available for search in this method: *CA-Condensates, Biological Abstracts* and *BioResearch Index, Medlars,* and *Chemical-Biological Activities.* This system was used at the 157th National American Chemical Society meeting in Minneapolis in 1969 and at the FASEB meeting in Atlantic City and the National Academy of Sciences this year. Other remote demonstrations have included a State Technical Services meeting in Texas and a group of companies in the Los Angeles area. Through the interactive nature of the system, the observers can see and understand within minutes how the search operates and the type of results that can be expected. When a typewriter terminal is used, the user carries away with him a list of the titles and bibliographic citations for his question.

The on-line system is used extensively in conjunction with the workshops as well. Once the basic principles of profile coding have been introduced to the workshop participants, they can operate the terminal and enter their questions directly. By reviewing the answers which are obtained, the question can be modified immediately by adding or deleting terms or changing the search strategy; and the question can then be researched. Thus, the user can quickly refine his question, seeing in a few minutes the results of many searches. Comparable searches using the standard batch search system would take days instead of minutes.

Overall, the goal of the educational program at the UGA center has been to get the services used as widely as possible. A more formal educational program undoubtedly could be designed, but it is doubtful that it would reach as large an audience or provide the same range of practical experience. The major support for the program comes from the Computer Center which is located administratively under the University's vice president for research. The program is considered an extension to the traditional library services, which are normally available to the faculty, students, and staff. And, as a side benefit, the same services are also made available on a cost recovery basis to other universities, government agencies, and industrial organizations who do not have or do not wish to develop the necessary computer facilities and support personnel.

ACKNOWLEDGMENT

The educational activities reported in the work were supported in part by the National Science Foundation under Grant GN-851.

LITERATURE CITED

(1) Caruso, D. E., "Tutorial Programs for Operation of On-Line Systems," *J. Chem. Doc.*, **10**, 98–105 (1970).
(2) Maizell, R. E., "Continuing Education in Technical Information Services," *ASCRAL*, 152 G015.
(3) Herner, S., "Experimental Course in Information Gathering for Scientists and Engineers," *J. Chem. Doc.*, **9**, 99–102 (1969).
(4) Rowlett, R. J., Jr., "Training Chemists in the Use of Chemical Abstracts Services," *J. Chem. Doc.*, **9**, 103–6 (1969).
(5) Jahoda, G., "Introduction to Symposium on Training Chemists in the Use of Chemical Literature," *J. Chem. Doc.*, **9**, 91 (1969).
(6) Fanta, P. E., and S. I. Miller, "Modern Techniques in Chemical Information. A Graduate-Undergraduate Course," *J. Chem. Doc.*, **11**, 98 (1971).
(7) Gelberg, A., "A Workshop in Chemical Substructure Information Handling Systems," presented before the Divisions of Chemical Education and Chemical Literature, 160th National Meeting, ACS, Chicago, Ill., September 14, 1970.
(8) Caughman, M. C., and J. B. Ayers, "Computer-based Information Retrieval: An Exercise for the Chemical Literature Class," presented at the 21st Southeastern Regional Meeting, ACS, Richmond, Va., November 6, 1969.