

J. S. KIDD

Toward Cost-Effective Procedures in On-Line Bibliographic Searches

With clues provided from descriptions of courses offered by five professors at the Cranfield Institute of Technology, on-line searches of a variety of machine-readable bibliographic files were performed. When the end-product was a "packet" of highly selected, full-text documents, the service was only coolly received. When the end-product was a moderately selective, reformatted bibliography, the response was very positive. Possible economic and ethical implications are considered.

THE IMMEDIATE OBJECTIVE OF THE PRESENT STUDY was to test the feasibility of a new form of information service. The service itself was designed to cope with several problems simultaneously. One such problem is the hidden-cost aspect of computerized information retrieval. Hidden costs are those that are imposed upon the user in the form of time expenditures in completing the information search and retrieval transaction. These costs can involve such elements as the time employed in the clarification of the user's needs and goals, the specification of search terms, the development of search strategy, the selection of items to retrieve from among those identified by the search, and the initial perusal and secondary selection for intensive study of items from those actually delivered. One goal was to

formulate a service which would minimize the time-cost to the user generated by components of the information retrieval process.

A second problem is that of user acceptance of computerized, and particularly, on-line information retrieval systems.¹ It is an instance of the general problem of the acceptance of innovation. Of course, this is a problem only to those who are advocates of this particular innovation or to those who are students of the processes of diffusion as such. However, for the sake of the argument, let us assume the stance of advocates of on-line bibliographic systems.

Having taken that stance, one would, in general, wish that acceptance would be rapid and widespread among important constituencies. One such constituency would be teachers and researchers in academic settings. In particular, a most appropriate target for the acceptance of innovation is those that have relatively high attributed status.² These individuals are more inclined to accept innovations and are also likely to have the power to influence others.

It is also worth noting that these are the same individuals whose time is

J. S. Kidd is Acting Dean, College of Library and Information Services, University of Maryland, College Park. The research reported here was supported in part by the National Science Foundation (Grant No. SIS 74-14127). The author gratefully acknowledges the assistance of the library staff at the Cranfield Institute of Technology.

priced the highest in European and North American academic communities. This factor of specific costs per unit time links the first problem to the second and brings into focus a set of design specifications for the information service. These specifications may be summarized as follows:

1. The service should employ and display the utilization of computerized, on-line bibliographic search capabilities.
2. The target audience should be comprised of senior academics.
3. The service should be unobtrusive, that is, undemanding of the user in the sense of generating minimal time demands upon individual beneficiaries.
4. The service should encompass and integrate as many procedural components as possible from need identification to document delivery; that is, it should be more than a mechanism for the production of retrospective bibliographies.

APPROACH

Within the framework of the design specifications there remains considerable latitude with respect to the precise mode of implementation. At the initial stages of development, the determination of the precise mode was necessarily arbitrary. Thus, for example, the on-line system used for preliminary test purposes was the Lockheed DIALOG system. It might have easily been some other, but DIALOG was conveniently available.

Similarly, a wide range of choice was available with respect to the identification of individuals as representatives of the target audience. Again, convenience was a factor. Individuals were selected from the teaching staff of the engineering faculties of the Cranfield Institute of Technology.³ More particularly, those chosen had the extra attribute of being short-course directors for courses

scheduled within the time frame of the study.

Short-course directors were chosen because they appeared to constitute a particularly amenable sub-set of the total audience. For instance, they have an acute need to be up-to-date and organized with respect to a particular content area as part of their efforts in course preparation and presentation. In addition, they must publicly and clearly define the content area in the form of a course description because the course description is the main promotional device used to recruit prospective students.

GENERAL SERVICE PROCEDURES

The initial steps involved in providing an unobtrusive information service to short-course directors were as follows:

1. Examine the course description; identify key descriptor-type terms; infer the conceptual relationships within these terms (e.g., central terms, modifiers, etc.); and formulate a tentative search strategy.
2. Conduct the search using all potentially productive files; adjust the strategy as necessary while on-line to provide a bibliography of at least ten and at most one hundred items.
3. Conduct a relevance assessment of the final bibliographic product (i.e., the computer printout) against the criteria imputed from the short-course description; delete non-relevant items.
4. Re-edit the resultant list to eliminate items suspected to be of low value (for example, extensive series of quarterly progress reports on the same project).
5. Obtain the selected documents.
6. Review the documents obtained for relevance and prospective value; eliminate marginal items.
7. Package the remaining items in a convenient form; provide a cover

explanation and re-formatted bibliographic listing; deliver to client.

8. After a suitable interval, interview the client with respect to his or her evaluation of the product.

RESULTS

The results are extracted from a sample of five instances which are treated as preliminary case studies.

Case A

The course title for this case was "Environmental Impact of Transport Systems." The key terms initially selected were *Environment*, *Environmental*, *Pollution*, *Noise Pollution*, *Intrusion*, *Severance*, *Transportation*, *Transport System*, *Transport Systems*, *Cost Benefit Analysis*, and *Cost Benefit*. The on-line transaction revealed, however, that only the most generic terms were useful. The final search cycle used only *Environment*, *Environmental*, and the three variants of *Transportation*.

Three files in the DIALOG system were searched; INSPEC (Physics), COMPENDEX, and NTIS. A total of 135 items were identified. A rigorous, not to say ruthless, screening for relevance produced a yield of twenty-five items. This very low precision factor was very largely a consequence of three incidental influences which tended to inflate the initial catch: first, the catch included a large series of studies on the ambient-state atmospheric environment, which studies coincidentally were sponsored by the U.S. Department of Transportation; second, people working on air traffic control problems use the term "environment" very loosely (i.e., terminal environment to mean air terminal vicinity); finally, NTIS is required to index all environmental impact statements produced by local authorities planning road-building projects, and these produced a voluminous chaff-factor of over twenty titles.

However, it should also be noted that

in addition to substantial deletion of non-relevant items, there was also an attempt to work toward a compact core, a set of documents that would have maximum value for minimum bulk. Such a core, it was hoped, would be economical for both the system and the user.

The twenty-five items selected were ordered for delivery from the available library resources. Some twenty-two items were on hand within a two-week period. These documents were screened on the basis of a review of the full text. The final "core" consisted of eight documents which were packaged and delivered to the short-course director.

After a suitable interval, the short-course director was interviewed to obtain evaluative feedback. Questions concerning the value of the materials, per se, and the utility of the packaging concept were broached. The general reaction was unenthusiastic. The principal criticism was that the materials were not particularly current. (Imprint dates were: 1971, two items; 1972, one item; 1973, two items; and 1974, three items; delivery was made in summer 1975.)

Case B

The target course for the second case study was "Energy Conservation and the Environment." The search terms derived from the course description were *Energy*, *Energy Resources*, *Fuel*, *Pollution* (with air, water, environment, chemical, thermal, and radioactive), and *Pollution Control*. In the actual search, these reduced to *Fuel*, *Pollution*, *Environment*, *Environmental*, and *Engineering*.

The search was run in the same time frame as the first search so the files searched and general approach remained the same. A total of 102 items were identified and recorded in the form of a full-citation printout. Fourteen were selected for actual retrieval. Again, local government impact state-

ments contaminated the catch, and the weeding was ruthless.

When the full text items were obtained, five were selected for packaging and delivery to the short-course director.

Feedback was again very cool. The short-course director indicated he had not seen some of the items before and that all items were relevant but that there was nothing in the set that was particularly useful.

While only a very small sample had been tried, the message was clear. The pruning process was being taken too far. Too much was being taken on by the service provider. The client was being deprived of his or her options.

Inasmuch as the document retrieval, full-text review, and package preparation were all high-cost steps and given the clues provided by the client evaluation, it was decided to back off considerably from the original scheme and to try a more conservative approach in which the client would be presented with a moderately trimmed bibliography rather than a "core" package of documents.

Case C

The name of the course in this instance was "Tribology." The search terms taken from the course description were *Lubrication*, *Lubricants* (liquid and solid), *Bearings*, *Friction*, *Testing*, and *Instrumentation*. In the actual search, it was found that *Lubrication* and *Testing* were sufficient to identify the required subject.

A total of fifty-three items were detected in three files, ISMEC, Engineering Index, and NTIS. A review of the printout indicated that eleven were not relevant and there was one duplicate (the same paper given at two symposia). Thus, there were forty-one hits or a precision factor of over 77 percent.

The forty-one citations were sorted into alphabetical order by author and retyped in standard GPO format. This

list was presented to the client.

As a cross-check on the prior procedure, a sub-set of eight titles was chosen as a simulated "core." The assumption was that the client would possibly choose a small number of the titles to be retrieved in full text and thus the client's selection could be compared with the service provider's selection in an unbiased manner.

Indeed, in less than a week, the list was returned with items ticked for acquisition. The number of hits, however, was seventeen. Four of the eight selected by the information provider had been selected by the client: a score no better than chance.

A note accompanied the returned list which expressed enthusiasm for the scope of the content coverage. Examination of the non-selected items revealed no discernible or consistent standard pattern. (For example, imprint date was not a pronounced factor, etc.) Clearly, however, the client had his own personal criteria which were valid in his framework.

Case D

The title of the course in this case was "Small Digital Computers." The search approach used in this case was different from that used in the other cases. The subject area was the application of small-sized computers to industrial processes. The course included consideration of computer arithmetic, computer organization, programming, peripheral equipment, and computers in systems context. All this appeared somewhat vast, and it was anticipated that the resultant bibliography even after screening would run into the high hundreds or even thousands of titles.

The approach taken was to emphasize the *teaching* of small-computer technology to mature students. Thus, the first search attempt was made via DIALOG on the ERIC file.

This proved a profitless exercise. It

was not possible to make a proper distinction, using ERIC, between teaching about computers and using computers as teaching machines.

A secondary search was made using the Psychological Abstracts file. This file did not contain any relevant items that could be detected.

Finally, a tertiary search was made using the Engineering Index file. Here we had just one hit, but it was perfect. The full citation is entered as follows:

Meyer, John D. *Minicomputers and their Real-Time Industrial Applications*. Texas A & M University, College Station, Texas, March, 1973; 100 p. Final Rept. USAMC Prof. U7422.

(Abstract: The paper presents a brief introduction to minicomputers and their real-time industrial applications. It considers how the subject might be taught to graduate-level industrial engineering students. The author covers minicomputer hardware, software, and peripheral devices.)

Notification of the existence of this item was passed to the client forthwith; and, not surprisingly, he was quite gratified and ordered its delivery. He had not been aware of this document before.

Case E

The title of the course in this instance was "British Transport and Society: 1800-2000." Potential key terms from the course description included *Transport, Transportation, Plan, Planning, Social, Political, History, Future, Futuristic*, and *U.K.* and all its variants. As might be expected, this was a cumbersome set of terms. The search was initiated using only the first four terms and slotting "Social" in and out depending on the file.

The first file searched was NTIS. Of the five variants of *Transport Plan* connected by *or* and four variants of *Social*, four items were identified, one of which was relevant. The next file was Engineering Index. With the use of the

"Transport Plan" set only, nine items were identified of which six were relevant. The ISMEC file was tried next using the same broad approach. Four items were identified, none of which were relevant. Finally, the ABI (business data) file was tried, again with the simple, broad approach. Nine items were identified of which four were judged relevant.

In summary, four files were searched producing an aggregate of twenty-six items of which eleven were real "hits." This is an interesting case of the persistence required in coping with a semi-soft content area. It also demonstrates some of the adaptations possible in the real-time mode.

In any case, the eleven citations were retyped and passed on to the client. It might be noted that only three items contained content explicitly concerning U.K. The others were either general or U.S.-oriented.

The client's response was again enthusiastic. He specifically remarked on the "accuracy" of the content coverage with respect to what he characterized as a "vague requirement" specification. He remarked that several of the items were already known to him but others were both new and potentially of notable value. He remarked that all items were relevant. He already had access to most of the eleven items listed, but he asked that two be obtained specifically for his retention. One of these had specific British content—the other was general in orientation.

DISCUSSION

An immediate disclaimer needs to be entered with respect to the strength of the conclusions that can be drawn from five particular cases. Obviously, one emerges with only the most primitive of working hypotheses: hypotheses that need more extensive testing across a fair range of conditions and, ideally, with more intensive controls.

The point, however, was to make a start, and this was done. We now know somewhat more than we did before. Specifically, for example, it appears to be quite feasible to specify acute user needs and to formulate workable bibliographic search strategies on the basis of the clues provided by such user-produced materials as course descriptions.

The advantages are clear and manifold. The procedure is low-cost both to the user and to the provider. It is susceptible to corrective feedback control albeit the feedback is somewhat delayed. The feedback delay, however, is compensated for to a degree by the fact that the need specification and search can take place well in advance of the point in time at which the acute information need reaches its peak intensity. Thus, the procedure seems to be suitably "anticipatory."

We also know more about what not to do. It seems increasingly unlikely, for example, that it is either feasible or desirable to deliver "packages" of actual documents on an unsolicited basis to an unprepared client.

However, the response was generally positive, not to say enthusiastic, to the bibliographic list as opposed to the document package. Thus, one could conclude that an unobtrusive retrospective search service is feasible and attractive to clients when the delivered product is the bibliography, as such.

Many questions are left unanswered. We do not know yet whether exposure to the service sample will encourage acceptance and support of on-line bibliographic systems by the specific individual recipients—much less the entire body of senior teaching faculty of the institution.

At worst, some potential individual sources of active or passive resistance to this innovation have probably been eliminated. At best, the individual beneficiaries will become active sources of support and will spread the word, vol-

untarily and informally, among their immediate peers.

Against a background of reasonably rapid diffusion of acceptance among junior staff and advanced students,⁴ even the worst case variant could have a significant positive effect.

A second question would involve the explication of the differences between the unobtrusive service described here and profile-controlled, selective dissemination services. This point is particularly salient because the form of the deliverable product observed to be most acceptable to users is a document list, something which does not differ much from the deliverable product of SDI services. Thus, in some ways, SDI-type services and unobtrusive services such as those described could be considered as competitors for the same niche in the general service environment.

While no specific test was made, it is possible to speculate usefully about the potential advantages of an unobtrusive-type service. In this regard, the unobtrusive service is chronologically discrete and therefore generates computer costs only infrequently. Likewise, the user is not constrained to review and revise his or her profile as a repetitive chore. Thus, both system costs and costs to the user are likely to be lower for the same size user population, compared to SDI services.⁵

Probably the more crucial unanswered question is one which has ethical overtones. It might be stated thus: Accepting the potential economic net gain to the client derivable from an unobtrusive approach, to what extent should the service provider feel empowered to intervene in manipulating the contents of the bibliographic list which is delivered to the client? In other words, can the ethical service provider act to select (i.e., delete) items for the final bibliographic listing—and, if so, on what grounds can the selection be made?⁶

It is probably not prudent to expect

that the first part of the question can be even partly resolved by research. The ultimate resolution of that aspect will rest with the ongoing interaction of the value systems of real clients with real service providers. However, if the answer is a qualified affirmative to some form of selective intervention, the second part of the question becomes a difficult but worthwhile challenge for research. For example, the case study experience hints at the prospect that at least some editorial cleansing is appropriate in the form of re-typing the entries in standard citation format—eliminating both the typographical and incidental content peculiarities of the usual computer printout.

In the cases reported using the revised procedure, only items were eliminated that seemed of very marginal relevance. A few items were dropped that seemed to be low-yield filler material—such as the quarterly progress reports mentioned above.

There was a temptation, which was resisted, to delete items that would pose great difficulties of acquisition. For ex-

ample, items published in remote areas were considered askance but were retained. Other problematic items were ones that seemed out-of-date, superficial in treatment of the topic, or repetitious of other items.

By and large, these items were retained because the grounds for selection seemed speculative. However, the fact is that had any been deleted, there would have been no way for the client to know. In this regard, there was no implicit or explicit claim that the lists were comprehensive. A rather high-handed selection procedure *could* have been adopted without the client being aware of it.

Other professions are noted for their prescriptive approach: the service provider is alleged to know more about what the client needs than the client. This is not the traditional approach of the information service professions and probably cannot be in the future—but the nature of the service described here raises some interesting points for debate and some important researchable problems.

REFERENCES

1. Joan M. Maier, "The Scientist Versus Machine Search Service," *Special Libraries* 65: 180-88 (April 1974).
2. Everett M. Rogers and F. Floy Shoemaker, *Communication of Innovations*, 2d ed. (New York: Free Press, 1971), p.185-86.
3. The Cranfield Institute of Technology is a post-graduate institution specializing in advanced work in engineering and business administration. In addition to programs leading to the M.Sc., M.B.A., and Ph.D. degrees, it provides a wide range of refresher courses for technologists in industry. The student enrollment is small (i.e., less than 700) and predominantly residential at the rural self-contained campus forty miles north of London. It is famous as the site of the Cranfield Project work on relevance assessment. See C. W. Cleverdon, "The Cranfield Hypothesis," *Library Quarterly* 35:121-24 (April 1965).
4. Elizabeth McCauley, *Results of the Cranfield Experience Within the O.S.T.I. Short-Term, On-Line Information Retrieval Project*, final report (Cranfield: Cranfield Institute of Technology, 1975).
5. Louis W. Stern, C. Samuel Craig, Anthony J. LaGreca, and Gerald J. Lazorik, "Promotion of Information Services: An Evaluation of Alternative Approaches," *Journal of the American Society for Information Science* 24:171-79 (May-June 1973).
6. Recent considerations of ethical problems have focused on the aspect of dealing with politically sensitive issues in an ethical manner. See, for example, Bernard Vavrek, "Ethics for Reference Librarians," *RQ* 12: 56-58 (Fall 1972).

Check List For

REPRINTS RELEASED IN '76-'77

Nachbin, L.: THE HAAR INTEGRAL	Orig. ed. 1965	168 pp.	9.75
Nachbin, L.: TOPOLOGY AND ORDER	Orig. ed. 1965	128 pp.	In Prep.
Nachbin, L.: ELEMENTS OF APPROXIMATION THEORY	Orig. ed. 1967	131 pp.	In Prep.
New, J. F. H.: OLIVER CROMWELL	Orig. ed. 1972	124 pp.	3.95
O'Flaherty, F.: THE CHEMISTRY AND TECHNOLOGY OF LEATHER	Orig. ed. 1965	448 pp.	In Prep.
Osserman, R.: TWO-DIMENSIONAL CALCULUS	Orig. ed. 1968	469 pp.	In Prep.
Perrin, D. D.: ORGANIC COMPLEXING REAGENTS: Structure, Behavior, and Application to Inorganic Analysis	Orig. ed. 1964	376 pp.	In Prep.
Phillips, E. L.: DAY TO DAY ANXIETY MANAGEMENT	Orig. ed. 1977		In Prep.
Poole, P. A.: THE UNITED STATES & INDOCHINA FROM FOR TO NIXON	Orig. ed. 1973	255 pp.	In Prep.
Rankama, K.: PROGRESS IN ISOTOPE GEOLOGY	Orig. ed. 1963	705 pp.	32.50
Rappaport, A.: THE MONROE DOCTRINE	Orig. ed. 1964	122 pp.	3.95
Reich, L. & Schindler, A.: POLYMERIZATION BY ORGANOMETALLIC COMPOUNDS	Orig. ed. 1966	750 pp.	In Prep.
Restle, F.: PSYCHOLOGY OF JUDGEMENT AND CHOICE	Orig. ed. 1961	247 pp.	In Prep.
Rickers-Ovsiankina, M. A.: RORSCHACH PSYCHOLOGY	Orig. ed. 1960	656 pp.	After pub. 22.50
Rooney, J. R.: AUTOPSY OF THE HORSE: TECHNIQUE & INTERPRETATION	Orig. ed. 1970	164 pp.	14.95
Rooney, J. R.: BIOMECHANICS OF LAMENESS IN THE HORSE	Orig. ed. 1969	271 pp.	16.50
Rosenfield, I.: HOSPITAL ARCHITECTURE; Integrated Components	Orig. ed. 1971	336 pp.	29.50
Rothstein, M. F.: GUIDE TO THE DESIGN OF REAL-TIME SYSTEMS	Orig. ed. 1970	253 pp.	In Prep.
Rouse, H.: ADVANCED MECHANICS OF FLUIDS	Orig. ed. 1959	458 pp.	17.50
Sansone, G.: ORTHOGONAL FUNCTIONS Vol. 9	Orig. ed. 1959	425 pp.	18.00
Schultz, R. J.: THE LANGUAGE OF FRACTURES	Orig. ed. 1972	408 pp.	16.50
Schwartz, A. M.: SURFACE ACTIVE AGENTS AND DETERGENTS	Orig. ed. 1958	855 pp.	32.50
Scott, W. T.: THE PHYSICS OF ELECTRICITY & MAGNETISM	2nd ed. 1966	720 pp.	17.50
Scott, A.: ACTIVE AND NONLINEAR WAVE PROPAGATION IN ELECTRONICS	Orig. ed. 1970	340 pp.	In Prep.
Sellin, T. S.: DELINQUENCY: SELECTED STUDIES	Orig. ed. 1969	161 pp.	9.00
Shockley, Wm.: ELECTRONS AND HOLES IN SEMICONDUCTORS	Orig. ed. 1950	581 pp.	In Prep.
Skilling, H. H.: TEACHING—Engineering, Science, Mathematics	Abridged ed. 1976		In Prep.
Slyter, E. M.: OPTICAL METHODS IN BIOLOGY	Orig. ed. 1970	765 pp.	44.50
Smith, C. R.: MECHANICS OF SECONDARY OIL RECOVERY	Orig. ed. 1966	504 pp.	In Prep.
Snead, R. E.: ATLAS OF WORLD PHYSICAL FEATURES	Orig. ed. 1972	158 pp.	14.50
Snell, F. D. & Snell, C. T.: COLORIMETRIC METHODS OF ANALYSIS Vol. IV—Organic Compounds (Complete set available)	Orig. ed. 1954	676 pp.	27.00
Solomon, D. H.: THE CHEMISTRY OF ORGANIC FILM FORMERS	2nd Revised ed. 1975	370 pp.	18.50
Thiensch, H.: DEFECTS & FAILURES IN PRESSURE VESSELS AND PIPING	Orig. ed. 1965	416 pp.	18.50
Thorp, E. O.: ELEMENTARY PROBABILITY	Orig. ed. 1966	162 pp.	In Prep.
Todd, C. D.: ZENER AND AVALANCHE DIODES	Orig. ed. 1970	278 pp.	In Prep.
Torrance, E. P.: GUIDING CREATIVE TALENT	Orig. ed. 1962	290 pp.	11.50
Wagner, A. F. & Folkers, K.: VITAMINS & COENZYMES	Orig. ed. 1964	532 pp.	27.50
Weitz, H.: BEHAVIOR CHANGE THROUGH GUIDANCE	Orig. ed. 1964	225 pp.	11.50
Whelan, M. J., et al.: ELECTRON MICROSCOPY OF THIN CRYSTALS	Orig. ed. 1967	558 pp.	22.50
Wilcox, C. H.: ASYMPTOTIC SOLUTIONS OF DIFFERENTIAL EQUATIONS AND THEIR APPLICATIONS	Orig. ed. 1964	259 pp.	In Prep.
Wood, C. T.: PHILIP THE FAIR AND BONIFACE VIII, State vs. Papacy	2nd ed. 1971	124 pp.	3.95
Wren, M. C.: WESTERN IMPACT UPON TSARIST RUSSIA	Orig. ed. 1971	272 pp.	4.50
Wright, F. B.: RURAL WATER SUPPLY & SANITATION	2nd ed. 1956	347 pp.	In Prep.
Wymore, A. W.: A MATHEMATICAL THEORY OF SYSTEMS ENGINEERING The Elements	Orig. ed. 1967	361 pp.	In Prep.
Youden, W. J.: STATISTICAL METHODS FOR CHEMISTS	Orig. ed. 1951	136 pp.	In Prep.

ROBERT E. KRIEGER publishing co. inc.

P. O. Box 542
Huntington, N.Y. 11743

A Quarter Century experience in Technical and Scientific Publications