

ASIDIC newsletter

No. 47, Fall 1983

ASSOCIATION OF INFORMATION AND DISSEMINATION CENTERS.

FALL MEETING HELD IN PHILADELPHIA CYRIL CLEVERDON GIVES KEYNOTE ADDRESS DAVID GROOMS ELECTED PRESIDENT

The Fall, 1983 meeting of ASIDIC was held in Philadelphia, PA on September 19-20 at the Franklin Plaza Hotel. The topic for the meeting was the concepts for pricing and marketing online information services. It followed the usual format: a keynote address, several shorter presentations, discussion groups and reports, a business meeting, and, of course, an Monday evening event. This meeting was also highlighted by the election of officers; David Grooms of the U.S. Patent and Trademark Office (formerly of NTIS) was elected President. Other officers were also elected; a complete list with names and telephone numbers is appended to this Newsletter.

The technical program was highlighted by a keynote address by Cyril Cleverdon who, as many know, has long been active in the information field. He began his address by identifying some key people in the development of the information science profession and then presented a paper entitled "Optimizing Convenient Online Access to Bibliographic Databases." A summary of Cleverdon's paper and the other presentations appears in this Newsletter.

The Monday evening entertainment featured a Philadelphia-style block party at the headquarters of BIOSIS (the host for the meeting), with typical Philadelphia foods.

Committee Reports

At the Fall business meeting, attendees were told that ASIDIC continues to enjoy financial health. A report from departing Program Committee Chairman, Marvin Wilson, was received. Bob Lormond, new chairman of the Program Committee, asked the membership to submit ideas for future meetings. The Planning Committee was asked to review the By-Laws and recommend changes three months in advance of the next meeting.

Future Meetings

The next ASIDIC meeting will be held March 18-20, 1984, at the Monteleone Hotel in New Orleans, LA. Host will be Marjorie Hlava of Access Innovations. The Fall, 1984 meeting will be in Washington, DC,

hosted by Lois Granick of the American Psychological Association. The Spring, 1985 meeting will be somewhere in Florida, and the Fall meeting of that year will again be in Boston (another clambake!).

New Publications

As Newsletter Editor, I receive many announcements of new publications or databases of potential interest to ASIDIC members. In this and future Newsletters, I intend to list a few of these. The list for this Newsletter is as follows:

- Information Intelligence, Inc., Phoenix, AZ has begun production of ONLINE JOBLINE, a database featuring employment opportunities in the library automation and online field. They also have announced *Online Libraries and Microcomputers*, a newsletter.
- The National Bureau of Standards has issued their report on the workshop on a materials property data system held in November, 1982 at Fairfield Glade, TN (see ASIDIC Newsletter No. 45). Copies of the report are available from Dr. John Rumble at NBS.
- INSPEC has announced the creation of Section D of their database, devoted to information technology. The hard-copy equivalent is the publication *IT Focus*.
- Information Market Indicators, headed by Martha Williams, has produced the first audit of the online database industry.
- The Association of American Publishers (AAP) is holding a one-day symposium on December 12, in New York. Its topic is "New Technologies in Libraries--Impact on Publishers." Rita Lerner, ASIDIC Past-President, is arranging the program.

New Members

ASIDIC welcomes the following new members. The name of the representative is listed following the address and telephone number.

U.S. Patent and Trademark Office
CP2-9C27
Washington, DC 20231
(703)-557-7338
David W. Grooms

Burrelle's Clipping Service
75 E. Northfield Rd.
Livingston, NJ 07410
(201)-992-6600
Mary E. Edsall-Tracy

Industrial Technology Institute
2901 Baxter Rd.
Ann Arbor, MI 48109
(313)-763-9273
Maurita P. Holland

NewsNet, Inc.
945 Haverford Rd.
Bryn Mawr, PA 19010
(215)-527-8030
(800)-345-1301
Gary Reibsamen

ABC/CLIO Information Service
2040 Alameda Padre Serra
Santa Barbara, CA 93103
(805)-963-4221
Marie Ensign

Lembaga Pendidikan Komputer Indonesia Amerika
Jalan Wijaya X No. 3, Kebayoran Baru
Jakarta, Indonesia
773436
Jusuf Randy

(Address Correction)

Fachinformationszentrum Energy, Physik, Mathematik GmbH, welcomed as a new ASIDIC member in the last Newsletter, has advised us of their full name and address:

Fachinformationszentrum Energie, Physik, Mathematik GmbH
D-7514 Eggenstein-Leopoldshafen 2
W. Germany

Fall Meeting Summary

Keynote Speech

Cyril Cleverdon, Cranfield Institute of Technology

Cleverdon began with an introductory reminiscence, in which he identified people who have had a major impact on the information scene since the 1950's. These are:

- Ranganthan, who developed facet classification and whose ideas still influence many of today's thesauri.
- Robert Fairthorne, whose papers on classification were years ahead of their time. He applied some mathematical principles to classification theory. Cleverdon identified two of Fairthorne's papers, "Classification of Mathematics" and "Mathematics of Classification," as landmarks.
- Calvin Mooers, who indexed and collected Fairthorne's papers. He developed Zator coding and has the distinction of being the first person to use the term "information retrieval*."
- Saul and Mary Herner were the first people to apply librarianship as a business in contrast to a profession. They established an independent consulting company based on their training as librarians.
- Eugene Garfield used modern technology to reproduce title pages as a current awareness device and developed the techniques of citation indexing. He recognized the potential of using computers for indexing and information retrieval.
- Mortimer Taube analyzed a catalog of some 50,000 items and indexed it using a vocabulary of 7,000 to 8,000 terms. From this came the uniterm system of coordinate indexing.

Cleverdon then proceeded to deliver his paper, "Optimizing Convenient Access to Online Bibliographic Databases." He began by observing that many discussions focus on cost-effectiveness of online systems; little attention is given to convenience of the users. (Cost-effectiveness was defined as the level of performance obtained for the lowest possible cost.) Many of today's systems, in their effort to be as cost-effective as possible, are overly hostile to the users. Although they use the trappings of modern technology, they are inefficient and

* Throughout his talk, Cleverdon stressed that what many call "information retrieval" today is really the retrieval of citations leading the user to the desired information.

expensive. The demand for online systems emerged before anyone knew of all the problems involved. Unfortunately, many of the original problems remain today. Because of the hostile nature of online systems, intermediaries have emerged; Cleverdon believes that the intermediary will continue to be needed, even as microcomputers proliferate.

Research has shown that when two different groups of people construct a thesaurus for a given subject and then index a set of documents, only 30% of the index terms will be the same. If they search for a subject in a database, 40% of the retrieved items will match, and if they assess the relevance of the search, there will be 60% agreement. Optimization of search systems is therefore difficult. The application of a controlled vocabulary can be justified for a printed index, but not in an online environment. A controlled vocabulary is intended to improve recall, but it lowers precision; results of studies therefore inevitably favor natural language. Abstracts are better for retrieval than indexing.

Cleverdon proposed an alternative to Boolean searching, which he termed inefficient. Boolean techniques divide the universe into two categories: those which match the query and those that do not. All citations in the matched set have equal probability of being relevant; output should be presented in decreasing order of relevance for convenience to the user. Cleverdon thus proposes that search systems use the quorum technique rather than the Boolean. Quorum searching has the following advantages:

- It is simpler for the user.
- The output is in logical order.
- It is more flexible than Boolean searching.
- It reduces search time.
- It is user friendly.

In a quorum search, the system looks for items with all of the desired terms present (like a Boolean AND). If it does not find any, it drops one of the terms and repeats the search. This process is repeated, dropping each term in turn, until a match is found. If none occurs, two terms are dropped, and so on. The first item found is, presumably, the most relevant, the second next most relevant, etc. The search can be stopped at any time when the desired levels of precision and recall have been reached. Quorum searching is most successful with a high level of indexing or when the abstracts can be searched free-text.

In contrast to quorum searching, consider the following steps needed for a Boolean search:

1. Decide the words to be searched.
2. Decide upon the database to be searched.
3. Translate the words into index terms.
4. Guess the terms likely to have been used by the indexer.
5. Prepare the Boolean statement.
6. Find the input instructions.
7. Enter the input correctly.
8. Repeat the above until the results are satisfactory, then examine the output to find the most relevant citations.

Using quorum techniques on a natural language file requires only steps 1, 2, 6, and 7, and then a decision on how much output to accept. The widespread use of Boolean searching explains the need for intermediaries; quorum searching is much more user friendly and easier to understand.

Another impediment to user convenience is the multiplicity of databases. The objective of 100% coverage is laudable but not necessary and not in the best interests of users. Duplication of papers in overlapping databases is wasteful of storage. Cleverdon proposed a single database covering all of science and technology, although he admitted that political and practical considerations would probably make it impractical.

Research has shown that there is a finite limit to the amount of information a user will read. A recent study found that if ten citations were judged relevant, only five were consulted, and if 100 were relevant, 28 were read. The amount of information gained per paper decreases as the number of papers read increases. It is rare for a user to read *everything* on a subject. The needs of most end users can usually be met by a few good papers. Such findings present strong arguments against the development of comprehensive databases. Instead, Cleverdon proposes one database covering some 400,000 papers per year, with records containing titles, bibliographic information, and abstracts. This database should be mounted on a system which supports natural language searching using quorum searching techniques. Cleverdon feels that such a system could be established easily if ten major scientific publishers agreed to support it, if costs to the user were reasonable, and if the system were user friendly.

New Trends in Information Delivery

Richard Kollin, ISI

With a note on the pitfalls of forecasting, Kollin outlined five trends or areas of controversy in the present online database industry.

1. There is a trend to increasingly friendly systems, of which there are three kinds:
 - **Forgiving** systems which allow the user to make errors in spacing, abbreviations, etc. They are capable of translating synonyms into the requirements of the system.
 - **Friendly** systems issue prompts to the user and provide help screens or menus. Examples of such systems are BRS/After Dark or Knowledge Index.
 - **Cuddly** systems learn from the user and store the user's interactions. They are responsive to the idiosyncrasies of individual users.

The controversy in this trend lies in the potential challenge to the intermediary. Kollin feels there will continue to be a need for information specialists, but their role will change to that of a consultant or expert.

2. Systems for delivering search output in a form more useful to the user are beginning to appear. The results may be ranked according to potential relevance, or a technique called "zooming" may be used. Zooming has been developed by the European Space Agency on their ESA/IRS system. The search output is ordered based on an inverted frequency list of elements (words, authors, etc.). Zooming can be thought of a "Visicalc" for online systems.
3. Keyboard aversion can cause impediments to information delivery. People who make decisions may regard keyboards as demeaning; most probably have not had to use keyboards in their careers. This phenomenon may disappear as the current generation familiar with microcomputers moves into management.
4. Multiple locations of a database give users problems. Kollin envisions a single "gateway" system providing access to many systems. End user access to online systems can be expected to burgeon when such a gateway becomes available. The database publishers will be able to reach a much larger market, and market duplication could be eliminated.
5. There is a conflict between personal computers and videotex-type (smart vs. dumb) terminals. In Europe and Canada, the trend seems to be towards the videotex terminal, with the intelligence resident in the mainframe. In the U.S. and Japan, the trend is

towards the microcomputer or intelligent terminal. These two approaches have implications for information system design. Kollin predicted that the videotex approach will prevail.

In conclusion, Kollin proposed the establishment of an "Edsel Museum of Not-so-Smart Inventions in Information Retrieval" which would contain optical coincidence cards, sorting needles, and micrographic libraries. He suggested that the museum be used to project future technology!

Online Searching: Costly or Cost-Effective? A Marketing Perspective

Harry Boyle, Chemical Abstracts Service

Boyle addressed the issues of why people buy online services, and are these services cost-effective. He pointed out that the customer perceives the value, which must be greater than or equal to the cost of acquiring the product. The customer's perception determines buying behavior.

Factors which influence perceptions of the value of information services include the evident (visible) costs, the uncertainty of a return on the investment (decisions are rarely attributed to information), and the possibility of deriving a long-term return from a short-term expense. Information may not be directly or immediately productive; therefore, the costs may be seen as an overhead expense. Overhead expenses are especially susceptible to cost cutting. Information may be used by only a few people; it may not have a quantifiable economic value.

Factors influencing the purchase decision include the need for information and the benefits expected from using it, the price of the product and the customer's ability to pay that price, and the element of risk involved in finding the information.

Perspectives on the value of information vary according to the viewpoint of the observer. The purchaser who authorizes payment may not understand the value of the service and may expect 100% recall and 100% precision. The searcher who formulates the strategy and uses the system has the best grasp of the costs and is in the best position to decide on the benefits. The consumer who uses the results may be unaware of the limitations, costs, and risks; they just want the solution to their problem. A single person may represent one or more of these views, complicating the market.

Some insights may be gained by considering price changes for online services and users' reactions to them. Price has a symbolic aspect; customers often associate high price with high value. There is also the implicit assumption that items paid for are items of value. So, for example, customers may think of the connect hour as the item of

value, rather than the information retrieved. Proposals for charging a flat fee per search have floundered because users are used to the connect hour method of charging. The connect hour as the charging basis has serious deficiencies. It leads to rushed searches, little interaction or browsing, and it tangles the roles of customers, suppliers, and vendors. Vendors are tempted to pad response times to increase connect times instead of pricing by the value of the information.

Suppliers and vendors must adopt a marketing orientation and emphasize what the customers want instead of mechanics or technology. Vendors need to realize that they do not know customer needs very well and should work towards understanding those needs better. The rapid growth in online searching implies cost-effectiveness, but value is based on the information itself. Prevailing pricing policies are counterproductive because they do not charge for the item of value--the information--but rather an artificial measure of value.

Selling to the End User Market

Deborah Silcox, Mead Data Corporation

Mead pioneered selling full text services to the end user market. They estimate they have trained about 600,000 users. Their revenues have grown from \$2 million in 1972 to an estimated \$100 million in 1983. Mead defines end users as those who need the information, i.e., anyone not a professional information specialist or intermediary. The end user market was chosen by Mead because it is much larger than the library market. End users may need education about the process of retrieving information; with intermediaries, this step is not necessary.

Mead began marketing its LEXIS system to the legal market because it is research-intensive--all the end users need information. They developed a system with a user friendly interface, fast response time, and a dedicated special-purpose terminal. The service was designed to be full text to allow it to compete with traditional hard copy products.

To provide such a service, a vendor needs a computer, communications network, terminals, software, training staff, support staff, and, of course, a database. Since the vendor provides everything, the vendor can control pricing. An up-front (or minimum usage) charge controls the number of users; Mead initially set this charge high because of a small training staff. By 1976, the marketing strategy was changed to appeal to smaller law firms; the minimum usage charge was dropped in favor of a "search unit" charge based on the amount of work the computer needed to do. Users paid a lot to search, little to browse, which encouraged longer connect times and more practice time. This scheme had the disadvantage that users found it hard to understand,

and it occasionally resulted in abnormally high charges for searches. An upper limit to the charge for a single search corrected this problem.

Mead's charging scheme has these advantages:

- Response times are fast.
- The search unit approximates the work performed by the system.
- More efficient users are rewarded with lower charges.

It has these disadvantages:

- It is difficult to understand.
- It is unpredictable and therefore unbudgetable.
- Charging back cannot be done until the bill arrives.
- Users tend to do complex searches to overcome the pricing scheme.

In summary, Silcox listed the elements of a good pricing scheme:

- It should be flexible. Small searches should be less expensive than large ones.
- It should be budgetable and predictable.
- It should allow for fast response times and system enhancements.
- Users should be encouraged to stay online.
- It should discourage direct comparison with the competition.
- Downloading should be allowed for, thereby discouraging illegal downloading.
- It should reflect the value of the information to the user.
- And it should be as simple as possible.

Obviously, no single structure can meet all these objectives.

Looking to the future, Mead sees no real problems with downloading because of the full text nature of its databases. Universal terminals will expand the end user market.

What Searches Should Cost and How to Sell Them
Janne Hunter, University of Pennsylvania

Librarians and information specialists often have to convince potential clients of the value of information. Library customers do not think of the library in terms of value--time saved, mistakes avoided, etc. Information is not necessarily used at the time it is obtained; requesters sometimes seek out documents a considerable time after the citations were retrieved. Then they do not remember where the information came from. For these reasons, we need to develop a way of valuing information and to make people aware of this value.

Intermediaries need to know the actual costs of information to be able to inform customers and justify prices. Why cannot vendors and producers publicize costs? Why is a database cheaper on one system than on another? Are some databases subsidizing others?

Connect time is not an appropriate charging measure because it discourages interactive searching. Some services charge more for 1200 baud access--as much as four times as for 300 baud. This is unfair because not all search time is spent in output. Currently, considerable attention is being paid to downloading and pricing for it. For one time use, there is no difference between paper output or machine-readable output; therefore, there should be no surcharge for downloading.

Hunter suggested that a possible solution to the dilemma of connect time charging would be to charge for output characters, with some fields possibly being surcharged. Royalties should be stated separately so that system users can be informed.

Marketing and Pricing for Full Text End User Services

Debbie Hull, BRS, Inc.

The following technological advances are changing online services and their pricing strategies:

- Higher baud rates are rapidly becoming the standard. Most of BRS's customers now access the system at 1200 baud, and a few are at 9600. Higher baud rates reduce connect times and hence revenues.
- The popularity of microcomputers means that more and more users can download data and store it cheaply. As time progresses, micros will become irresistible for many users. The market for online services will therefore increase, but downloading could have serious implications because users will create their own databases and search them on their micros, reducing revenues to online vendors.
- Full text databases are growing, and vendors will have to find efficient ways of creating them. Manual input is labor-intensive and expensive. Full text files may be obtained as by-products of the photocomposition process, but conversion to an online database is

difficult. Publishers need to coordinate printing with the online product. Storage costs are coming down, making full text databases more attractive. The same software can be used for searching these files, but enhanced features for printing are necessary. BRS has developed some of these.

- Software is being developed to help end users do searches. Many systems feature menu-driven interfaces as well as all the power of the original software. Examples are Knowledge Index and BRS/After Dark. End users want lots of features; they rapidly become tired of simplistic approaches. They do not read documentation but want on-line help, escapes to help messages, and preprogrammed function keys. They often need help setting up their micro and getting started.

Future technologies impacting online services include the laser disk, cheap local mass storage, digital encoding of data, analog information to recreate images, and graphics capabilities.

Different approaches are needed to market services to different markets. End users use full text databases to get answers to questions; intermediaries tend to use them to survey a topic. End users want the *answers* online, not abstracts. They are motivated to search full text databases by content, and they want *current* information. Online tables of contents are attractive; the lack of tables and graphics in full text databases is a drawback. The librarian's role will change to a teacher of strategies, or a specialist on hardware.

Experience at BRS suggests that pricing strategies should be hardware independent so that vendors will not delay necessary upgrades because the costs cannot be passed on to users. They should also be simple and understandable; end users want single invoices. And pricing should allow a fairer approach to 300 vs. 1200 baud searching. Some possible approaches to pricing are:

- A connect hour charge allowing only online printing. The royalty would be high enough to take care of the reprint business. No off-line printing would be allowed.
- A charge for printing certain fields online. Database producers like a structure which charges for the number of full text paragraphs printed online. The present BRS system cannot handle this.
- A per page charge, such as that used for the ACS Journal Database.
- The flat fee "all you can eat" approach. BRS will offer this as an option for their Kirk-Othmer database produced by John Wiley.
- A combination, including a monthly or annual fee for a minimum amount of searching with excesses charged at a per hour rate.

- A CPU-based charge such as used by Mead or I. P. Sharp. This may cause problems because vendors would have to re-educate their current users, and prediction of search costs would be difficult.

In the future, Hull sees the telecommunications networks charging by the number of characters transmitted. It is possible that the traditional approach where vendors pass the telecommunications charges through to users may change. Some telecommunications services may become vendors, but they lack the software that the present vendors have spent many years developing.

Discussion Groups

The following questions were posed to the discussion groups:

1. What role, if any, should database users have in database production?
2. How can the producer/vendor help the search intermediary market database services?
3. What recommendations can be made regarding user-friendly protocols, pricing, or training as a marketing tool?
4. What would increased government regulation of scientific-technical information transfer do to present end user marketing strategies?

Most groups thought that user input was most useful regarding content and coverage of databases, but of little help in database design. Considerable interest was shown in the requirements of end users. Many thought that intermediaries should market database services in general, and not limit themselves to one database of a limited number of them. Concern was expressed about the lack of formal marketing training in many librarianship courses.

In the discussion on pricing, the group seemed willing to admit that prices should cover costs. The intermediary viewpoint seemed to prevail: vendors should provide information on the cost breakdown. It was suggested that Cleverdon's quorum techniques, if offered by a vendor, might provide an opportunity to explore different pricing schemes. One person suggested that a database is no different from a book; the purchaser does not care how the price is determined, and the seller has no interest in how the book is used. A single purchase charge allows the user to make unlimited use of the information in the book.

The discussion on training covered ground that had been gone over before. Training is good, but expensive. It is difficult because database producers and vendors are not acquainted with the needs of their customers and do not know what to teach them.

In the discussion on government regulation, some attendees suggested that monopolistic practices and restraint of trade in the database industry is a danger. The government is also showing signs of restricting the flow of information to certain countries.

Acknowledgement

I am greatly indebted to Judy McQueen, Senior Consultant, Information Systems Consultants, Inc., and Betty Unruh, Dialog Information Systems, for their very thorough notes of the meeting.

ASIDIC Officers, 1983-1984

President

David W. Grooms
U.S. Patent and Trademark Office
CP2-9C27
Washington, DC 20231
(703)-557-7338

Member-at-Large

Dennis Auld
Data Courier, Inc.
620 S. Fifth St.
Louisville, KY 40202
(502)-582-4111
(800)-626-2823

Immediate Past Present

Rita G. Lerner
American Institute of Physics
335 E. 45th St.
New York, NY 10017
(212)-661-9404

Member-at-Large

Marjorie M. K. Hlava
Access Innovations, .Inc.
P.O. Box 40130
Albuquerque, NM 87196
(505)-265-3591

Secretary/Treasurer

Member-at-Large

Taissa Kusma
American Mathematical Society
P.O. Box 6248
Providence, RI 02940
(401)-272-9500

Member-at-Large

Robert Lormand
Lawrence Livermore Laboratory
P.O. Box 5500
Livermore, CA 94550
(415)-422-5277

Committees

Committee	Chair
STANDARDS	Maureen Kelly
PROGRAM	Robert Lormand
PLANNING	J. Ron Smith
FINANCE	James D. Cape
PUBLICATIONS/NEWSLETTER	Donald T. Hawkins
MEMBERSHIP	Cathy Ferrere