INTRODUCTION

Selective Dissemination of Information (SDI) is a new and rapidly developing field. The concept was originally set forth by Hans Peter Luhn in 1958. As described by Luhn, one part of a larger idea, the business intelligence system, was Selective Dissemination of Information. SDI involves the use of the computer to select from a flow of new documents, those of interest to each of a number of users. This process may be thought of as the inverse of information retrieval. In information retrieval, a user precipitates a search of a file of documents. In SDI a document precipitates the search of a standing file of user interests. SDI has been called "current awareness" since the attempt is to keep the user aware of current developments. This function has been traditional with those few really excellent librarians and executive staff assistants. SDI is a mechanization of this function.

The concept described by Mr. Luhn was first implemented in 1959. At that time in Yorktown Heights, New York, an IBM 650 Data Processing System, together with other card machines, reproduction equipment and human operators, processed a small flow of documents against the interest profiles of some 30-odd users. This system has subsequently been called "SDI 1." In 1960 a second system, SDI 2, evolved from the original one. SDI 2 is the first system designed and documented so that it may be installed remotely. In 1961, documentation for the SDI 2 System was completed and the first public announcement of a documented SDI System was made on July 11. Implementation started on a third system, SDI 3, early in the year. Although SDI 3 was in partial operation during the last part of 1961, complete debug and documentation was not completed until 1962. During 1961, several other systems, all of which will be designated by their location names, became operational. These included SDI Owego, at the IBM Federal Systems Laboratory in Owego, New York; Current Awareness Service, a System at the Technical Information Center at General Electric's Evandale installation, and a manual current awareness service available at cost to all United States Citizens at the Office of Technical Services of the United States Department of Commerce in Washington, D. C.

More systems became operational in 1962. The second system tested and documented for remote installation, SDI 3, became available through the IBM Data Processing Library for the IBM 1401 Data Processing System. The Poughkeepsie System went into operation near mid-year at the Technical Information Center of the IBM Data Systems Division in Poughkeepsie, New York. A fourth system from the Mohansic Group, SDI 4, has been in partial operation for nine months and documentation...
is well under way.\textsuperscript{11} The Douglas Aircraft Corporation in Santa Monica, California, has a system in an advanced state of debug.\textsuperscript{12, 13} Around the end of 1962, a second 1401 program became operational at IBM, Data Processing Division, Midwest Region, in Chicago, Illinois. Over the past four years, SDI has moved from a concept into a rapidly increasing number of system installations.

Implementation

Implementation difficulties for a system are often underestimated. This is in contrast to reduced operational cost and increased quality of service which are often overestimated. With SDI, the choice is whether (1) to use an available system, (2) modify an available system, or (3) to write your own new one. Because of the uncertainty, implementation cost is hard to estimate. Quality is even more difficult to estimate. Everyone seems to feel he is an expert on quality. There is disagreement in many cases. Present SDI systems involve computer programs, manual procedures and sometimes special equipment. In order of increasing difficulty, implementation may involve the installation of a well-documented, tried and true system which is in operation somewhere else; modification of manual procedures; obtaining special equipment; reprogramming or redesign.

Human skills available; experience of the personnel with SDI, Information Retrieval or related areas; and the number of other systems, procedures or constraints interacting with the new SDI installation all affect the effort required for implementation. Not only are a wide systems background, computer knowledge, and documentation experience valuable, but specialized knowledge with office machinery, industrial engineering, typography as well as psychology, sociology and organization theory often help. Programmers seem to be necessary for any type of installation. The more experience with data processing as contrasted with scientific programming the better, but any programming experience is better than none. Systems and procedures personnel are well-known in most organizations and are certainly advantageous for modifications or rewriting.

Experience with installations of documented SDI systems is limited. It was estimated that three calendar months and a total of three man months effort would be necessary to install an early SDI system.\textsuperscript{3} Programmers have been used in all cases. The time to get a SDI program through a monitor system or to fit in with other existing operating procedures has been quite variable. In one case the program assumed a particular load routine long in general use, but not in use in this installation.

Experience with combining and modifying existing systems is exemplified by Poughkeepsie. There, despite the fact that the programs had little, if any, documentation, one or two programmers fought through SDI, KWIC* and an IR program in a few months. The manual procedures were in flux for a longer period. The total system is still being modified and only parts are in operation. Owego was a rewrite from SDI 2 which took over a calendar year to get into operation. The program rewrite itself, from start to run, took about three months. Prestart systems work extended longer and, to my knowledge, the system is still rather weakly documented for remote installation\textsuperscript{18, 19} and is being integrated with KWIC.\textsuperscript{7} What might seem to be a relatively simple rewrite of SDI 2, SDI 3, required one person for a calendar year in a building being noisily rebuilt, although the programming and documentation was done by an experienced programmer who knew SDI and the machine.

The classic problem seems to be an underestimation of the amount of the programming required to rewrite and document. For experienced personnel, e.g. SDI 3, estimates seem to be low by a factor of four. For less experienced (with SDI) personnel perhaps six would be better, e.g., Poughkeepsie. It should be pointed out that certain phases can sometimes be estimated accurately, e.g., programming at Owego.

User Interests

Most user profiles (interests) have been obtained without any problem by blindly mailing a short form to the potential user.\textsuperscript{†} In three tests\textsuperscript{‡} some 65% of those contacted became users. Mass meetings of potential users have been used as well as blindly mailing longer

\textsuperscript{†} Key Words in Context, a machine prepared printed index.\textsuperscript{17}

\textsuperscript{‡} Pages 94-5.

From the collection of the Computer History Museum (www.computerhistory.org)
forms with either term dictionaries attached, e.g., Owego (modified ASTIA), or enclosing examples*** of indexed document items. Indirect methods have also been used to derive profiles from personnel or project information.*** Only with SDI 1 was a comparative study made and it had too small a sample to be conclusive.*** Each of these methods have been proven feasible. Further research is needed to define situations where one is preferable to another.

Adjustment of user profiles has been done largely at the user's instigation. At Mohansic, blanket mailings of current user profiles with change forms have been made to encourage users to make changes. Users have also been notified that they can make changes. The effectiveness of these measures is subject to doubt. The only known attempt to automatically update or adjust profiles based on user's responses was tried at Mohansic on SDI 1. The results were inconclusive. Manual attempts to suggest or arbitrarily make changes in user profiles based on various hypotheses have been made from time to time, usually without controls.

Although how to get new users to join and give the "best" possible profile seems to be a difficult theoretical problem, in practice there seems to be no difficulty. Experiments with automatic updating are in order but adequate user response histories seem to be necessary.

The number of users serviced by SDI systems now in operation has ranged from tens, to one to two thousands. Experience with larger groups is lacking although no new problems are anticipated. One problem, not initially anticipated, which increasing number of users has proven to be important, is that of address changes. These occur so frequently that not only must they be considered part of every normal run, but provision is necessary to change addresses between notification and hard copy order. As we shall see below (Abstracts and Notifications) this affects the notification itself.

*** Ibid., Page 6.

Documents

Document sources for SDI are usually defined by the application. The range of subject matter on which there is experience is quite wide, including science, engineering and management. There are no known cases of letters, memorandums, or picture annotations being processed although this has been proposed and no problems are anticipated. Document source has been shown to be a significant factor in response.14 Owego uses ASTIA documents predominately. Poughkeepsie uses internal IBM reports. Surveys of what users read15 or library usage could be used to determine what document sources to use for an SDI system. Most such data indicates a skew distribution of usage with a few highly used journals. It is assumed but not demonstrated that different types of users need different document sources. Experimentation in this area might influence the selection by professional journals of items to abstract. SDI provides a tool in this area through its response. SDI 4 and a revised Chicago system will allow exclusion of documents by source, e.g. need-to-know or excluding journals user subscribes to. Volumes of document items being processed in SDI systems run from tens to hundreds per day with experience upward lacking. Subscribing to a journal is not much of a problem, but getting on internal distribution lists is more difficult than one might expect. It cost the Mohansic group several man months of effort to locate internal sources of information and arrange to be added to these distribution lists.

Documents normally come to one location, are handled and numbered. Some SDI's integrate with library operations to various degrees. Owego uses the same numbering and hard copy reproduction procedures. Mohansic provides abstract sets and utilizes journals from the library. Douglas is partially integrated. Some work with IR systems, e.g., Evan-dale, Owego, Washington. Document numbering may be sequential as at Mohansic or by an internal code as at Owego and Poughkeepsie. Checking for duplication and series completeness is a normal library problem.

There seem to be few serious operational problems in this area. Studies are needed to test automatic procedures to analyze user responses and to vary the document source mix to maximize value functions. Little has been done to study the effect of frequency of mailings to the user.
Indexing and Decision

The primary decision method used in SDI has been a probability of interest estimate, \( i/d \), where \( i \) is the number of words identical in the two lists and \( d \) is the number in the document list.* The words are normally chosen by humans** from text and truncated to adjust for endings. This particular technique came from a programmer's misunderstanding in 1958 of H. P. Luhn's instructions. It proved to work and was therefore kept. The proposition is that probability of interest increases with \( i/d \). The \( i/d \) criterion may vary by document—Mohansic—or by user—Owego, Poughkeepsie and Chicago.0 A no-truncation fixed-dictionary system with a thesaurus (Owego, ASTIA thesaurus) has been used with \( i/d \). Experiments have been run at Mohansic, but not as yet reported, on truncation, 4-9 characters;* and depth of indexing, 1-26 keywords; as well as machine indexing from partial text by several methods. Conventional “Boolean” methods with keywords are used at Evandale. Chicago indexes by machine from the abstract using KWIC methods, i.e. dropping common words—A combination \( i/d \) and “Boolean” method with variable truncation is used.

A variety of indexing and decision procedures have been used. There needs to be work to compare the results under varying conditions. Relationships between SDI and IR need to be explored empirically in the indexing and decision area. Are they the same or different; if different, in what ways? More work needs to be done on the desirable amount of direct user control over the decision.

Abstracts and Notifications

The decision is made to notify the user of one or more items. What should the notification consist of? In one of the SDI tests, hard copies were sent directly to the user.† Users preferred a two-stage over a one-stage procedure: receive abstract notifications and be able to order hard copy instead of receiving the hard copy directly without intermediate control.‡ However, their reading habits seem to be considerably more effected by direct documents.*** With this exception, SDI systems have used two-stage procedures sending abstracts to users and allowing them to obtain hard copies, in some cases providing order forms. Abstracts have been compared to titles and the titles seem adequate for deciding which document to order, but abstracts are necessary partially to substitute for the document if the document is not available.14 Quality of abstracts has been discussed in theory. There is no known experimentation. SDI and IR, with appropriate response evaluation, would appear to be excellent vehicles for this exploration.

A number of forms for notifications have been discussed and tried. IBM cards have been the most frequent vehicles. The abstract is normally typed on a reproduction master and reproduced onto cards. Normal card stock works well with offset or stencil but spirit master runs are too short. Chicago and Owego use or will use the IBM 1403 Printer to print on continuous-form abstract cards directly. A machine record of the form number and the user and document would have to be kept so that when the card was returned the response could be mark sensed and the number read. Mark sense requires special pencils which deposit an electrically conductive mark, but new optical machines (e.g., the IBM 1418 Optical Character Reader and IBM 1428 Alphameric Optical Reader), allow standard No. 2 and No. 3 pencil marks. Machine (1403) printing of the form number in place of prepunches is possible with an odd font. These have yet to be tried for SDI but appear cheaper for high volumes.

Single (SDI 1) vs. multiple (all other) card systems have been under debate since 1959. This debate no doubt will continue. The notification should combine (1) the document abstract (preferably both 3 x 5” and IBM card size), (2) the user's address (3 or 4 lines for complete postal address which constantly changes), (3) the system return address, (4) questions regarding the document, (5) provision for the user's remotely made response to the questions, (6) the document number and (7) the user identification. The notifications

---

* See* Page 30; Page 9, 10; Page 8, 10.
** Education level doesn't seem to make any difference. See*.
† See*; ‡ Page 6.
‡ See; † Page 10;‡ Page 8-9.

*** See* Table IV.
should be in appropriate sequential order for mailing. If 5, 6 and 7 are not machinable on return, response handling for document hard copy orders and operating statistics must be manual, as in SDI 1. The abstract, 1, should be retainable by the user. A study in one organization shows 3 x 5 and IBM card sizes were the most frequently used media for this purpose even prior to SDI. The response, 5, is made at many remote uncontrollable locations.

The PORT-A-PUNCH® card has proven to provide a machine readable response. PORT-A-PUNCH is only now (February 15, 1963) becoming available in continuous forms, thus making machine (1403) printing of the abstracts on the PORT-A-PUNCH card or an attached form possible. Previously a bill feed attachment was necessary which slows the printer. Systems remain to be developed and tested based on bill feeds, optical reading and many other devices. When several notices go to each user at once, placing several cards together (or using a sheet of paper as at Evandale) might save handling expense and user exasperation. No existing system meets all of these requirements; each compromises to some extent. Considerable research is necessary before sufficient basic knowledge is obtained as to the relative worth of these various features.

Response, Reports and Hard Copy

SDI 2-4 require the user to respond on every notice. Other systems require responses under certain conditions (SDI 1, no response if negative) or never, i.e., just a notification. It is not known exactly what effects this has.

Responses and other records allow reports to the user, operators, management and research personnel. This is a largely undeveloped area even though some rudimentary reports are included in the SDI 2 and 3 systems. Feedback reports could be used to assist in updating user profiles, changing the document sources mix, adjusting the system sizes, changing indexing methods, and adjusting the cost vs. value balance. Randomly selected notices (SDI 1-4) allow the system selection performance to be compared to random selection as a base. This also allows miss items (which could have been selected by the system but were not) to be estimated statistically.

There have been various hard copy procedures. (1) Ignore the problem (SDI 2-4). (2) Refer the user to a library (SDI 2-4). (3) Shelve and pull (SDI 1-4). (4) Keep vellum and reproduce (SDI 2-4). (5) Use aperture cards and reproduce (initially at Owego). (6) Use reel microfilm at multiple locations (Poughkeepsie). Adequate analysis of cost and value are yet to be made. Most systems agree with the Mohansic survey, users want to be able to obtain hard copy.

Value-Cost

This is, in my opinion, the area with the largest potential for development. Available cost data is very limited and hard to interpret. Available value information is largely subjective. Dichotomy scales have been used in SDI, i.e., “of interest vs. not of interest.” It is my opinion that ordinal, and cardinal scales are needed if we hope to move SDI design from an art towards a science.

REFERENCES


