Design and Use of Information Systems

by

Kieran P. Broadbent
International Development Research Centre
Ottawa, Canada

Paper presented at the Second Summer Institute on Information Science
May 16 - 21, 1977

Institute of Library Science
University of the Philippines
Diliman, Quezon City

*The views expressed in this paper represent those of the author and not necessarily those of the Centre.
Design and Use of Information Systems

by K. P. Broadbent

ABSTRACT

This paper describes key concepts involved in the design and use of information systems. It deals with the processing, storage and retrieval of documents and emphasizes user requirements. The discussion attempts to elucidate some of the common constraints on the design of new systems. Only when the basic requirements of a system are known can an evaluation of its effectiveness be made.

Information retrieval is a major component of present day science and technology. It provides scientists and society generally with a glimpse into the past as well as a method to grasp quickly the results of day-to-day research. Information systems are the means by which the general principles and techniques governing the transfer of information are efficiently applied. Each system has to cope with a vast and growing bank of knowledge which has to be acquired, stored, shared and retrieved again and again. As the volume of documents in a given sphere expands the system must progressively develop to deal with it.

It is important to consider what we actually mean by such common terms as "information" and "retrieval". Formerly it was quite easy for a librarian to identify the type of information being dealt with since most information came in hard copy form. Today the same librarian must be able to handle a wide range of 'packages' from patents and standards to audio visual matter. The task, therefore, has become increasingly complex. Information in its modern sense, however, can include a wide range of
material from numerical data to raw research results. The basic fact, however, that should be understood at the outset is that 'information', by the very nature of the term, is meant to be used and not stored and forgotten. The information system is like an amoeba, continually growing, splitting off, dividing and growing again. The user might also regard it as a jigsaw puzzle, since he will spend some 90% of his time chasing 10% elusive information, like the missing piece in the jigsaw. It is the task of the system to make his task easier by locating the elusive data as rapidly and as cheaply as possible. It is now usual to consider systems design in the establishment of information services. We know, for instance, that information science, a relatively new concept in terms of the history of librarianship, is the study of the conceptual theories and practice of information flow and transfer. In other words, it is 'that discipline that investigates the properties and behavior of information, the forces governing the flow of information for optimum accessibility and usability.'

Documentation, on the other hand, is one of the component parts of information science concerned with the acquisition, storage and retrieval of documents. Information science has increasingly become a highly complex subject, inter-disciplinary in its academic approach especially in its response to new methods and developments.

Systems design, on the other hand, has only lately come to be a term linked in many librarians minds when information systems come up for discussion. Systems design is basically what the term implies: it is the basic configuration of the system. The systems analyst tries to make optimal use

---

of resources, both human and material, to make the system work effectively. The theoretical basis is mathematics, but in practice the systems designer will have to call on a whole combination of disciplines, logic and language (software) and engineering (hardware).

Information systems, as previously stated, concern the flow or exchange of information usually in the form of records of documents. We can, therefore, refer to this simply as 'input', 'throughput' and 'output'. In other words, the collection, processing, storage and retrieval of documents. The first hazard in systems analysis is defining the system. For instance, if the system is too specific it will fail to provide the full range of information users expect to receive. On the other hand, if this system is too broadly defined it will be nebulous in scope and in practice becomes hopelessly unworkable and costly. Systems design, therefore, is a critical factor in a modern information system.

The establishment of an information system requires three basic principles:

1. bibliographic control, by which is meant tracing relevant material for input into the system;

2. management control, which includes design and operation of the system in such a way that retrieval is efficient; and

3. cost control. The value of any system can be judged on whether data can be retrieved and disseminated from the file relatively cheaply.

With these three factors in mind, constraints are constantly being placed on the system which affect the general efficiency and overall performance to the point where a system will progress or fail. In order to supply the desired service a retrieval system will need as near perfect accuracy
as possible, accepting the fact that 100% coverage is an ideal, but rarely attained goal. The amount of irrelevant literature should, however, be kept to a minimum. No user wants to sift through masses of unjustified material. Therefore, the structure and design of the system is crucial to its long-term working. The basic structure may be represented by diagramatic form as a stylized input-output model, thus:

![Diagram](image)

**FIG. 1**

The double facing arrows in the diagram depict provision for additional file handling methods, as for instance by subject categories or accession number. To gain access to the file the user will usually require an intermediary. Here it is the librarian's job to frame the user's requests in such a way to get the system to react (retrieval procedure). Bibliographic control is the essential element here. The depth and broadness of acquired documents will constitute the file's major usefulness (user satisfaction). Obviously, in a large library this will go hand-in-hand with overall acquisitions policy. However, in order to achieve a plausible depth of coverage for an information system, it really becomes vital to ignore traditional ways of treating a document as a single item. The process of assigning a large number of
descriptors to account for the document content will also invariably introduce an intolerable degree of irrelevance into a system. This becomes obvious if two articles are selected on branches of the same discipline, e.g. marketing and trade, and descriptors are assigned to cover all important items of the contents; the resultant retrieval chaos becomes alarming. The sort and analyze stage is designed to obviate somewhat this kind of retrieval difficulty. Each item should be tagged and classified according to an agreed list of subject categories.

Large scale operation of this concept can perhaps best be illustrated with reference to the FAO/AGRIS input/output operation, where a large retrospective "master file at FAO is linked to AGRIS input." A computer interface programme converts input in AGRIS format to meet the earlier FAO format on file, separating input into two batches; (1) containing AGRIS records; and (2) complete records validated through FAO software. The end-product of batch one is monthly tape from which AGRINDEX is produced. Batch two up-dates the FAO master file producing a monthly "FAO documentation current bibliography (see fig. 2)."

An important part of the design process will focus on the actual storage of documents. In 1967, the costs of storage of a single document were estimated to be $50. The systems analyst, therefore, must devise the most efficient means of storage. Simple forms of storage can be by author or subject classification using UDC or other principles but greater utility

---


FIG. 2
will be available if an entry can be stored by multiple file. Such files can be organized differently. Basically, there are two types of mechanized retrieval systems: (1) the serial access and (2) what is usually referred to as direct access but is sometimes called the "inverted" file. In systems which use serial access, records are arranged in linear fashion. The major disadvantage of this method is that in order to retrieve all records of single term usually means searching the whole file. Moreover, in order to input a record located in mid-file, it is necessary to run half the file before the relevant record can be reached. This represents several hours expensive computer time. The inverted file concept uses disk storage in which documents located at any point can be rapidly retrieved. In the inverted file, data are located as physical units on a serial. Thus, to retrieve from storage on a direct access system, the computer can be instructed to read a specific index term on the inverted file, look up the record on the direct access linear file and retrieve the relevant document without the lengthy process of searching through the linear file.

The retrieval process is the point where the user meets the system. The user may for instance work through a "middleman" or intermediary, i.e. the librarian or information specialist, or may address the machine himself. This is pretty much the same case as in the normal library where a reader will have a choice between using the files, browsing through the stack or going through the reference librarian to expedite his request. The librarian with specialized knowledge of what is in the stack will usually re-frame a reader's requests and help him gain access to what he wants quicker than he would by searching unaided. In the same way an information officer or user specialist working on an automatic system will re-formulate a user's request. It is here that operation and performance of an information system is thoroughly tested since the effectiveness is
influenced by retrieval language and indexing. In effect this means that if hierarchical and other relationships are properly assigned and synonyms are displayed the search will easily run from broad categories to the appropriate specific term and document. Without a good thesaurus or "entry language" valuable time will be lost searching irrelevant literature.

A word here is needed to explain some of the terms used in retrieval language and indexing since it is one area of particular importance. A descriptor is a subject entry usually assigned to "describe" the main contents of a document. Usually it is a single word, e.g. "marketing", but it may also appear as a bound phrase, e.g. "food marketing". Descriptors may also be used to describe plant names or geographical entries. An identifier is an additional access point used, for instance, as a trade name, and is less structured than descriptors. A specifier is a device that permits precise description of a record. All descriptors are specifiers; however, a specifier is not necessarily a descriptor. This is to say one cannot usually use "pigmeat" as a descriptor but the term can be specified in the file by use of two descriptors: "pigs" and "meat". Entry terms are terms which appear in the vocabulary to provide access to the system. Therefore, all descriptors and specifiers are regarded as entry terms. But conversely, not all entry terms can be regarded as descriptors or specifiers. If "pigmeat" is not defined but is to be sectionalized under "meat", the entry term will be merely a lead-in term to the basic index. The basic index contains all descriptors and specifiers.

An information system may have many uses. The traditional system of storage based on library resources from which is offered a range of services must be considered a basic form. Abstracting, Selective Dissemination of Information (SDI) and current awareness services are also important
applications. Ultimately, however, it is the user which any system is
established to serve. It is the user who demands the system, it is the user
who modifies the system, and it is the user who finally causes the system
to close. The constant monitoring of user attitudes and requirements is,
therefore, very important. It therefore follows that the system should be
flexible enough to adapt to users' needs. The large amount of capital input
required to get modern systems operational today means that they must be
properly attuned to user requirements if they are to be effective and
worthwhile. Feedback (see fig. 1) is, therefore, essential.

The basic considerations should include full discussion of scope and
detail of subject coverage of the system. Limits should be set. The type of
literature to be indexed will also be a governing factor, especially if
abstracting or use of abstracts is to be contemplated. Moreover, if a
large amount of non-conventional literature or fugitive documents is to be
covered, different indexing priorities will be required compared, for
instance, to books. The size and growth of the file will impose self-
constraints; the larger the file the greater the degree of irrelevant
retrieval. As indicated above, an information system can be many-faceted,
i.e. can set out to be an abstracting service or a simple literature service,
etc. All services present special problems. Some present greater demands
on the human and material sources of the service than others. If the system
is user-orientated, the language employed has to be simple. Costs will
always be a key factor. Where resources are limited economy must always be
borne in mind. There are various devices available for monitoring costs, 4

4 See for instance: Broadbent, K.P. 'The cost-effectiveness of information
services in agriculture.' Seminar on Agricultural Information Systems,
Also Vickers, P. 'Ground rules for cost-effectiveness.' ASLIP Proc.
(1976) 617, 224-229.
description of which is not permitted within the scope of this paper. However, institutions contemplating a new system will want to have a clear idea of staff, equipment and operating costs. For a small operation, a useful device is to try to break down all costs into component parts from which can be determined basic costs. In any case, it is best to obtain an agreed definition of the objectives and scope of the service to enable a good pre-budgeting idea of eventual costs allowing also for areas of expansion. In addition a system, once operating, must, if it is to have any chance of being cost-effective, keep to the limits it has been designed for, otherwise costs will get hopelessly out of control. Maximum effort should go into providing output, rather than processing and refining input.

An information system should not in any way be a self-centred process. Therefore, a major aspect of the design is to identify documents. It is the concern of the operator to ensure availability. If an item in the system is on record but not available the user will quickly cease to have any faith in the system and without users a system must collapse. Therefore, a system must be able to meet the demands arising out of the retrieval process in the form of inter-library loans, requests for hard and photocopy. A truly extensive co-operative system for loans/photocopy/microfilm does not yet exist. It may not even be possible for the vast amount of non-conventional literature in various disciplines to be made fully available, but greater efforts are needed to see that users can actually read the copies requested. The existence of an inter-library loan system is significant. In many countries such a system does not exist. Individual libraries in such places cannot hope to match demand for new services. Co-operation at local, national and regional level, therefore, is an increasing priority issue.
Summary

This paper has discussed the general principles employed in the application of systems analysis to the design and development of an information system. It has pointed out that any systems' ability to cope with the various demands made on it by users will be dependent on the structured database, and thesaurus together with its ability to give adequate response to the different types of information it is required to deal with. The system should be attuned to local limitations and resources and adequate thought should also be given to availability of literature. The long-term operation will rest on user satisfaction.


Holm, B.E. 'Searching strategies and equipment.' American Documentation (1962) 13(1), 31-42.


