DOCUMENT RESUME

ED 033 732 LI 001 788

AUTHOR	Hayes, R. M.; Reilly, K. D.
TITLE	The Effect of Response Time Upon Utilization of an Information Retrieval System; A Simulation.
INSTITUTION	California Univ., Los Angeles. Inst. of Library Research.
Pub Date	Jun 67
Note	14p.; A paper presented to the Operations Research Society of America Annual Meeting
EDRS Price	EDRS Price MF-$0.25 HC-$0.80
Descriptors	*Information Services, Information Sources, *Information Systems, *Information Utilization, *Simulation, *Systems Analysis

Abstract

A simulation program is presented which considers the effect of response time upon the utilization of a multiple node information system. Specific examples of such systems are presented. They include: (1) information networks, (2) library systems, (3) multi-level files. The simulation characterizes users (in terms of their information requirements and expectations), information services, and nodes of locations at which these services are available. The effects of various allocations of resources to different levels of service, different locations, and different response times can be demonstrated, particularly in terms of the levels and kinds of utilization. (Author)

THE EFFECT OF RESPONSE TIME UPON UTILIZATION OF AN INFORMATION RETRIEVAL SYSTEM A SIMULATION

R. M. Hayes
K. D. Reilly

A paper presented to the Operations Research Society of America Annual Meeting

June 1, 1967

Institute of Library Research
University of California
Los Angeles, California

TABLE OF CONTENTS

ABSTRACT

A simulation program is presented which considers the effect of response time upon the utilization of a multiple node information system. Specific examples of such systems are presented. They include: (1) information networks, (2) library systems, (3) multi-level files. The simulation characterizes users (in terms of their information requirements and expectations), information services, and nodes or locations at which those services are available. The effects of various allocations of resources to different levels of service, different locations, and different response times can be demonstrated, particularly in terms of the levels and kinds of utilization.

INTRODUCTION

Modern organizations, whether business, government, or service, have become exceedingly complex and the costs of trial-and-error analysis to assess changes in their structure and function have become prohibitively costly. Accordingly, the methods of systems analysis and operations research are increasingly being called upon to assist managers and designers of large-scale systems in the evaluation of new concepts and designs. A leading technique of operations research is computer simulation. It offers a method for obtaining, in a few minutes time, information that, through experimentation, might take upwards of years to obtain.

A generalized simulation model is being developed by members of the staff of the Institute of Library Research, UCLA, in order to study problems of file organization. The word "file" in this context is very broad and includes: library systems; information retrieval systems, such as that described by C. Blunt; computer files and data banks; and memory systems in general. No single model can hope to be comprehensive enough to capture all the essential features of such an array of systems; furthermore, each kind of system has its own nuances. Nevertheless, there is some value in simultaneously studying several systems in hopes of discovering some new features they have in common and of gaining the insight that comes with study-in-breadth.

---

*Blunt, Charles R.: An Information Retrieval System Model, Defense Documentation Center: Defense Supply Agency. AD 623 590. October 1965.
Blunt, Charles R.: A General Model for Simulation of Information Storage and Retrieval Systems, Defense Documentation Center, Defense Supply Agency. AD 636 435. April 1966.

It is initially proposed to apply this model to evaluations of library network systems using data representing the characteristics of the multi-campus University of California. The initial attack will be on the point of user response to the network operation. Variations in service parameters, effected through variations in the allocation of resources in the network and through changes in the service facilities, will be studied in connection with the satisfaction to the user, measured by changes in the probability of use of the library system. The satisfaction to the user, in the model, depends to a large extent on the response time of the service (although there are other non-temporal factors, to be discussed later, which have some bearing on the user's behavior). Several alternative configurations and services in the system will be studied, and the attempt will be made to ascertain which possibilities of organization have some prospect for future development in real library networks. A full-scale model of all the operations, including all the points at which time delays arise in service. is beyond the scope of this description; our discussion therefore is limited to a "representative model".

REPRESENTATIVE MODEL

1. The Nature of the User

A simulation model of a library network is to be developed and its properties relative to its users is to be investigated. Associated with each user of the system will be a collection of parameters, among which are a basic few which will change and, in varying degree, reflect the user reaction to the system. The response of these changing parameters to changes in the system will be used to estimate advantages and disadvantages of possible alternative systems.

Each user of a library periodically develops needs for information. However, not every need is met with consideration by the user to use the library. A book on the shelf, a colleague down the hall, etc. may be regarded as the solution to the problem. The existence of these factors is taken into account in the model through a parameter, f_m (where the subscript m refers to the mth user of the set of users). When f_m is multiplied into the number of needs developed by the user the result is the number of needs in which the user considered use of the library. The user must also estimate the time within which his need for information must be met (the need-time), and compare it with the time on the basis of past experience, within which he believes the library can meet his need. The need-time may vary from a very short time to perhaps days or even weeks, depending on the nature of the need. Similarly, the expected time of service of the library is a function of his experience; accordingly, it may also vary widely. The comparison of these two times will determine whether or not the user will use the library. If the user decides to make a request and the library delivers the desired material to him in the required time, the user will be satisfied by that response; otherwise, he will not. In either of these alternative cases, changes can be expected in f_m and in the expected service time distribution. (In some cases, it may be possible to represent the expected time distribution by means of a single parameter. In such cases, only the mean value of the parameter would be changed. In other cases, both the mean value and some measure of dispersion could be involved. A satisfactory experience in library use could very well decrease the mean value of the expected time distribution. What the effects on the dispersion of the distribution might be is perhaps more difficult to say. Intuition seems to argue that satisfaction would be correlated not only with shorter mean service time but also with less dispersion.)

The other distributions associated with the user are those which are not likely to change very much, if at all as a result of system operation. One of these is, of course, the distribution of the need-time, mentioned above. Again, as with the expected time distribution, there may be one or more parameters associated with this distribution.

In summary, the user first develops a need for some desired material; if he considers using the library, he compares the need-time with the expected service time; if the latter is smaller than the former, he enters into the library system at some location. We have discussed everything in this sequence with the exception of development of the needs and the act of entry into the system.

2. The Nature of the Needs

The library system, as pictured in the model, is to be capable of handling requests for services of a variety of types: copies of books, newspapers on microfilm, speeches on tape, etc. It may also provide retrieval services, including individual items of data. There is a problem related to the level of detail in specifying the items of request; we shall discuss this aspect of the model more fully below when we discuss material availability.

The timing factors in the development of needs may be handled in either of two fashions. The first method is to assume a time unit, in which the user develops a number of needs. The second method involves using a continuous time scale and to allow individual needs to be generated at points along that scale. In either of these two methods, a random number is generated for the simulation model. In the first alternative, the random number is used to sample from some assumed probability distribution for the number of needs developed over period; in the second alternative, the random number is used to sample from a distribution of the time between needs. The latter method would be the approved method if all the time delays in the library model were to be studied. In particular, if one were interested in the time delays generated because materials are taken out of the library, one would probably be interested in temporal distribution of the needs to a higher degree of precision than would be provided in the case of using the number of needs developed over a fixed interval of time.

For most cases of library simulation with large numbers of users, many of which are similar in one fashion or another, it is advantageous to deal not with the individual user, but rather with collections of users.

Thus, in the model, the word user may be applied to all Professors of English at UCLA, or all Research Organic Chemists in local industry, etc. It is clear once again that many different kinds of grouping of individual users are possible. Note, in particular, in C. Blunt's study (loc. cit.) users are classified as their modes of entry into the system and whether they are qualified for certain rapid methods of delivery of the results of a computerized search scheme. The important point is that by giving this kind of a definition to user types, the number of user types is kept to a minimum. The group of users is, of course, composed of individuals and the parameters of the distribution for the group should, in principle, be traceable to the number of needs of individuals.

There is still one more issue that should be mentioned in connection with generation of the needs of users. In a library like the University of California Library, the primary purpose is to serve the faculty and students within the university system. Faculty and students of the state colleges and other schools, the members of the staff at other research organizations, and the general public are not the primary targets for service of the university library system. Admittedly, the university wants to serve such users in so far as such service does not interfere with its primary function, but it will place a lower priority on such service. There is more than one way to represent the fact that certain users' needs are not to be attended to with the same fervor as others. For example, a low priority value may be assigned to users of certain classes. Time delays in service for these users would be sufficiently large that use of the library would not be expected to be too delightful a prospect. Hence, these users would not use the library as often as others. Another way to accomplish a similar result but with somewhat different consequences is to assign to each user-class a special parameter which is to be multiplied into the number of needs that are generated by these users. The resulting number would then be the number of needs developed by these users which are relevant to the library system. Such a parameter provides the library with the option of refusing service to certain classes of users (perhaps, for certain types of requests) even though the service provided to these users is so very excellent that left to the users' determination, the user would utilize the library for all of his needs. When we multiply the number of needs by some other parameter, which reflects what we might call the importance of the user, we have then a relative need--a need relative to the library system--a need that the library will attend to.

The level or status of the individual corresponds to assigning a priority scheme to the users. Certain pre-emptive powers may also be appropriate to users with a certain status. Status can be of many different types. One type, suggested by R. S. Taylor and C. E. Hieber*, is to divide users into seven levels: Faculty and staff; graduate students; the four undergraduate years; unspecified undergraduates. Of course, one could seek a somewhat simpler scheme, perhaps, be merging freshmen and by sophomores; [illegible] lower- and upper-division students and doing away with the unspecified undergraduates.

The establishment of the pattern of needs, both in time and with respect to the type of materials desired, can be achieved experimentally. It is doubtful that library users have very precise notions of the service times according to the nature of the request. However, it can be assumed that the expectations of the users will to some degree reflect the actual service time distribution; it is of interest to trace the lag time in and extent of re-adjustment of the users' expectations when a change occurs in the system.

*Taylor, Robert S., and Caroline E. Hieber: Manual for the Analysis of Library Systems. Bethlem, Pa., Lehigh University Center for Information Science, 1965.

3. Location, Status, Subject Area

Both the generation of the users' needs and the response of the system to requests are affected by the services requested, the location from which they are available, and the locations of the user. For example, libraries are often set up according to fields of study, and such special consideration achieves a better service time for the particular group of users which is geographically closest and interested in those subject areas.

The subject area of the users is again a matter of choice. Various specialty lists divide science into hundreds of different fields of major concentration. The National Academy of Sciences-National Research Council has kept track of fields in which doctorates can be earned. Their system provides a list which has the potential of covering all the numbers from 000-999. For the purposes of a simulation model, however, it would be well to have a smaller number of user-group types. Taylor and Hieber (loc. cit.) have offered a six category list which, not unexpectedly, provides broad classifications, roughly comparable to the divisions of the university into colleges: arts and sciences; business administration; engineering; education. Besides these, categories are provided for administrative people and those in special curricula.

The location of the users in the model means the points of access to the system. Included are both the points at which the user may enter the system as well as the points at which service is provided. For example, a user may be permitted to phone in his request from his office, but may have to pick up the material at some specified pick-up location. Cost factors depend most heavily on the allowed methods of presenting a request and the means materials are provided by the system. These, in turn, depend probably on the type of communications equipment, the nature of the stored data, the distance materials have to be physically transported, etc.

4. The Response of the System

The user appears at some point of access to the system, provides the system with his status or priority level as well as his requests. In general, when the user establishes these credentials the library personnel will invoke some search strategy. For example, a request by a professor of the history of the Middle East will lead to one kind of strategy whereas a request by a professor of physics would lead to another. Basic to the search strategy is the balance between "distance" (in terms of the time it takes to get materials from one location to another) and the probability that the given material is at the location. Imagine a search for material on affairs of the Middle East. The user enters the system at some location. To simplify the discussion, let us assume he enters the main library at one of the campuses of the University of California. A search will be made at that location to determine whether the material is there. If there is no success, the library personnel will have to make an inquiry on another campus to determine if some other campus has the desired material. Suppose now that only UCLA has the material and that the user is at a campus such as Davis. If the personnel at Davis began to contact the libraries nearest to them, when perhaps they know fully well that in the area of Middle Eastern affairs, the library most likely to have the material is UCLA's, they may spend a great deal of unnecessary search time.

Material availability at any location can be expressed in a variety of ways. Perhaps the simplest manner in which to do so is to assign for each material at each location a number (between 0-1) which represents the probability the given material is the location. We can, in principle, measure this probability by keeping track of the percentage of the time that each material is in the library. There are certain difficulties in this, as with materials that are very popular when they are first available, and then are no longer popular. But such effects as this can be compensated for, if they are known. When we do represent material availability in this manner, we allow the possibility of the same material to be out of the same library more than one time, which is an impossibility. We mentioned above that we were going to represent a group or class of material as the unit of materials. If this is so, then to state that the same kind of material is "out of the library" more than one time is to state that the same kind of material has been requested by more than one user. Hence, there may be no really serious drawback in this representation.

A second representation is to keep account of the material items, by assigning to each material that leaves the library a certain time period in which to return to the library. This assigned time, would, of course, have to be a random variable, since users do not always keep books. An "occupation" number (1) could be assigned to all materials when they are in the library. When the material is out, a zero can be assigned to it. Part of the search strategy in this kind of picture would be to see whether material that was out was due to come back or could be recalled etc. in order to satisfy the user's need. This method of representation seems to be most useful when the individual material items are being treated. We have already intimated that any model that would attempt to take into account all the individual items would become unwieldy. Yet this is necessary if we are to characterize the circulation system of the library very accurately.

5. Evaluation of Results

The inital study will be related to the users' needs and how they are met by various organizations of the system; changes in the system will be studied next. To take account of the change in the users' attitudes about the library, we have the parameters of consideration, f_m, and the parameters of the expected time distribution. f_m is restricted to the range 0,1 and, for that reason is a convenient parameter to deal with. The mean value of the expected time of service will also provide useful information. Improvement of the system response should lead to an increased f_m and a decreased mean value of the expected service time. In many ways, the increase in f_m and the decrease in the mean value of the expected time of service will reflect the same consequence. It may be important to accumulate certain other statistics also. For example, it might be worthwhile finding out if dissatisfied users are from related fields of specialization. It may be useful to know what the expected service distributions are for groups of satisfied users as compared to groups of dissatisfied users.