

Literature Growth and the Retrieval System in Scholarly Communication

Bert Boyce

Assistant Professor
University of Missouri

ABSTRACT

A series of simple models indicates that literature growth is a prime objective of literature communications systems. The literature communication system is clearly successful in the sciences, and trivia and redundancy are a useful part of its operations. If one treats the literature and its behavior as physically observable phenomena one can base selection and subject analysis on objective criteria.

A strong argument could be made for not burdening the literature of information science with yet another model of a retrieval system and its environment. I believe, however, that the formulation presented here casts some new light on the purpose

of such a system, and leads to some conclusions on the nature of the growth of scholarly literature. These conclusions tend to refute the pollution arguments against publication, some of which have advocated the end of inducements for publication¹ altogether.

It would seem that the most useful view of scholarly literature is as a communication channel between scholars. This is a function it clearly serves, although there may well be other functions we may wish to take into account. This view would give us the picture in figure 1, a simplification, for a special case, of the classic Shannon and Weaver Model of a communication system².

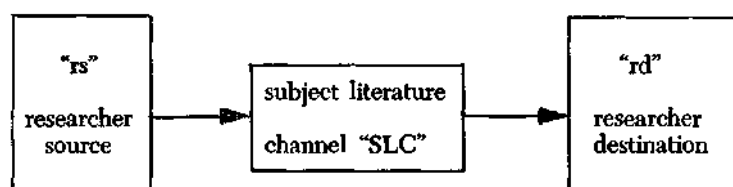


Figure 1.

The literature communication channel

Such a model, while accurate, is certainly inadequate to represent the scholarly literature system. First of all, the message sent through the literature is a broadcast one, a message from one to many, from the point of view of the source. From the point of view of the destination, it is one message for that destination selected out of a mass of messages in the channel from multiple sources. Thus it seems

the model should take into account not only the single researcher but that group of researchers which constitute the contributors to the subject. Figure 2 attempts to do this as well as to indicate that the literature of a subject is not the only channel of communication between the researcher and his peers.

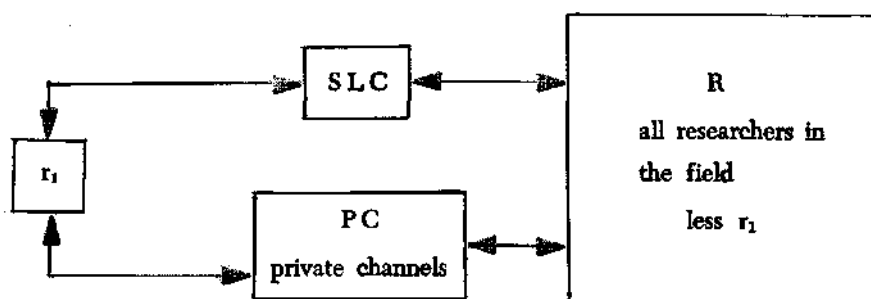


Fig. 2

Multichannel Scholarly Communication

The flow in the channels is now in both directions and the single researcher may be considered either as a source or a destination.

We are now in a position to observe one of the interesting results of this communication system. The subject literature channel does not clear itself after a transmission. Instead to accommodate the next message it grows. According to Price's³ figures for scientific literature it grows exponentially.

In fact, if we observe the results of the operation of the system the growth of the scholarly literature appears to be its reason for being. The scholar having selected messages from the channels available to him, and incorporating them with new observations drawn from outside the system, produces new messages which are incorporated into the scholarly literature.

At first glance it may seem somewhat strange to declare that the growth of the literature is the purpose of the scholarly communication system even if it is surely its result. It would seem that the growth of knowledge is the real concern of scholarship. I would support such a conclusion. When, however, one reads Ziman⁴ one may well conclude

that there is little difference between the two. One can strongly argue that information that is not in the public domain, is at best unconfirmed second class knowledge. Once material has passed through the filter of the publisher and reviewer system, and having been published, is exposed to the criticism of the scholarly community it gains considerable status as a contribution to knowledge. This clearly does not mean that every published paper has equal value. There is considerable trivia and redundancy in the scholarly literature. This, however, is a poor argument for slowing the growth of the literature. The rewards and incentives attached to scholarly publication are necessary despite the apparent pollution.

One can not expect the new researcher to leap in full armor from Jove's brow. Few scholars would care to have their reputations judged on their first efforts. Redundancy and triviality are costs in the development of scholars well worth paying.

If we remember Shannon⁵ we will agree that redundancy is not such bad thing in a communication system. It is in fact of considerable assistance in the operation of the Information retrieval system which now appears in our model in figure 3.

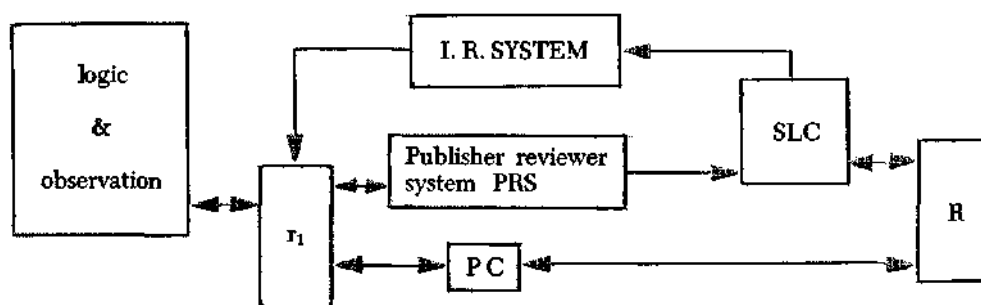


Fig. 3

The Information Retrieval system, like the publisher — reviewer system is a filter in the literature communication system. It stands between the scholar as destination and the literature channel just as the publisher - reviewer system separates the scholar as source from this channel.

Its purpose is to choose those messages which will be of value to the scholar in his work and reject the rest. Figure 4 shows its operation in greater detail. The main flow (shown by the solid lines) indicates the movement of documents through the double filtering process.

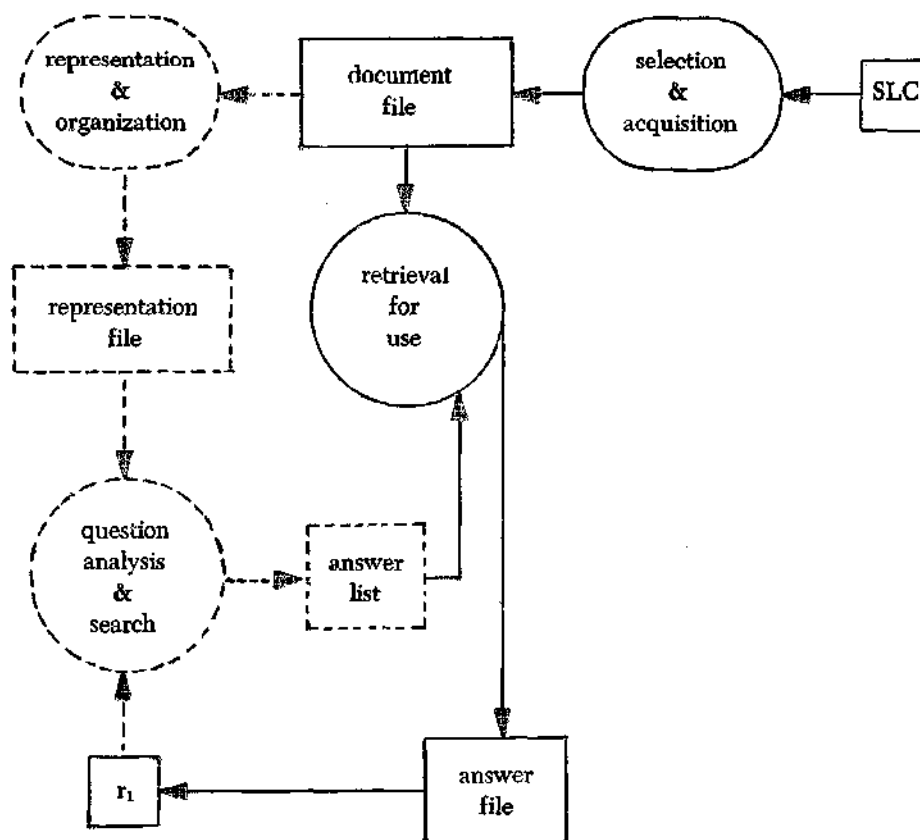


Figure 4

The information retrieval subsystem

They are selected from the literature on some regular basis, retrieved for use through a further selection process, and presented to the user. The first stage selects from the whole literature of the subject according to broad guidelines. The second stage filters on the basis of specifications. There is also a control flow (shown by the broken lines) which controls the operations of the second stage of the filter. The control processes may be carried out by the researcher himself informally, or may be handled by a formal organization. These retrieval processes have traditionally been carried out in a rather subjective manner. However, this does not seem to be necessary. If one assumes, as our model indicates, that the growth of the literature is the system goal, objective criteria suggest themselves. If one can identify important papers by measuring their effect on the growth of the literature of the subject, (by the number of citations made to them, perhaps) one can determine in which journals these papers most frequently occur. Thus the primary stage of the filter need not use subjective criteria. One

need hardly cite the wide literature on citation ranking and the use of Bradford's law of scatter. Neither is it necessary for the second stage to utilize subjective classes of documents. The classes can be formulated on the basis of the relation of these documents to one another in terms of common citations, common authorship, common authors in citation, common journals in citation or even common words occurring with high frequency in their texts.

A search method using such automatic classification techniques has been described by Goffman⁸. A review of the major classes of methods for creating objective classes of documents is provided by Yu⁷. There appears to be no absolute need for subjective subject analysis. If one can focus attention upon the subject literature as a physical entity with objectively observable characteristics and use these characteristics to stimulate the growth of the literature one has a practical and efficient view of a retrieval system which avoids a great many problems of currently apparent significance.

BIBLIOGRAPHY

- 1 — KENDIG, Frank. Drowning in data. *Saturday Review / Up Front*, August 1973, p. 26
- 2 - SHANNON, Claude E. & WEAVER, Warren. *The Mathematical theory of communication*. Urbana, Univ. of Illinois Press, 1949.
- 3 - PRICE, Derek J. de Solla. *Little Science, Big Science*. New York, Columbia Univ. Press, 1965. 118 p.
- 4 — ZIMAN, John M. *Public Knowledge*. London, Cambridge University Press, 1968. 154 p.
- 5 - SHANNON, C. E. op. cit.
- 6 - GOFFMAN, William. An Indirect method of information retrieval. *Information Storage and Retrieval*, 4(4): December 1968.
- 7 — YU, Clement T. A Clustering algorithm based on user queries. *Journal of the American Society for Information Science*, 25(4): 218-26, July/Aug. 1974.

RESUMO

Uma série de modelos indica que o crescimento da literatura é um dos objetivos primordiais dos sistemas de comunicação da literatura. Estes são claramente bem sucedidos nas ciências; o material não relevante e a redundância são componentes úteis de suas operações. Considerando-se o tratamento da literatura e seu comportamento como fenômenos fisicamente observáveis, pode-se basear a seleção e a análise de assuntos em critérios objetivos.