Introduction:

Why Analyze Bibliographies?

H. R. SIMON

IT IS becoming increasingly apparent that for the continued development of information studies it is above all necessary to investigate the paths along which scholarly information is moving. Since the scholarly exertions of antiquity, the flow of information has possessed vast reservoirs in the form of bibliographies. These represent a concentration of information such as may excellently serve, from the viewpoint of the statistical method, as representative samples for the investigation of the information flow as a whole. Bibliographies enable us to develop model formulations which may ultimately give rise to a comprehensive theory of information.

Bibliographies are worth investigating under carefully and methodically defined conditions. Two types of investigations could prove particularly helpful:

1. Investigations from the perspective of general information theory—the results of these inquiries would be of value to information specialists of all kinds. Classed in this area could be Martyn's article in this issue which gives an evaluation of the general situation regarding secondary literature; and Brookes's article which provides a detailed exposition of the methods of numerical analysis of bibliographies. An article was planned on the use of computer-stored bibliographies as sources for future developments in information and library science. Unfortunately this article did not materialize; however, an earlier *Library Trends* article by Malin describing the *Science Citation Index*®[1] might be helpful to the reader interested in this area.
2. Investigations from specialized viewpoints—the results of these analyses should reveal the current situation as well as trends in individual

H. R. Simon is Archivist and Lecturer for the Library, Zoological Institute of the University of Heidelberg, Germany.

areas of research. They are important for scholars in a great variety of specialized areas as well as for administrators in scholarship. In this issue the articles by Thompson on the humanities and Bottle on scientific bibliographies represent investigations from specialized viewpoints.

These two approaches to analyzing bibliographies, which naturally represent only crude classifications, must be kept in view when deciding to set up projects for the analysis of bibliographies. Lancaster has listed criteria important for the information scholar.[2] In addition to the well-known and frequently discussed factors of coverage, recall and precision, Lancaster distinguishes that of novelty. Novelty is a measure of the degree to which a bibliography brings to the user's attention references to articles that are new to him (i.e., articles that he was not aware of before seeing the bibliography). A novelty ratio[3] may be expressed in the form:

$$\frac{\text{Number of articles that are relevant to the user and new to him}}{\text{Number of articles that are relevant to the user}}$$

With the two criteria of novelty and costs, it may exactly be determined, above and beyond the general informational process, how a bibliography is rated from the viewpoint of the user. The fact, therefore, that there are two quite decisively important questions to be considered in the practice of information transmission may easily be shown by means of appropriate investigations.[4]

All the mentioned forms of analysis of bibliographies primarily concern information transmission. The conveying of information, however, takes place within one specialized context, within one certain field of research.[5] In view of this, bibliographical analysis is of importance to the scholar himself, although it can offer a valuable stimulus to the information specialist as well. To be more specific, we can ask the following questions, for example, in order to determine the information flow in a specialized area by means of analysis of its bibliographies:

1. Which journals cover the world's literature about _____ (art, physics, or whatever)?
2. In which countries are these journals published?
3. In what languages is the literature about _____ written?
4. How much is written in each subject field of _____, where is it published, and in what language?

In the above series of questions, 1, 2, and 3 are equally important for

information scholars and specialists, while question 4 has value primarily for the specialist and secondarily for the information scholar. Investigation of all the questions is possible in numerous narrowly or broadly specialized areas such as physics,[6] biology of mammals,[7] or cardiovascular medicine.[8] It would be a meritorious task to undertake as complete as possible a bibliographical listing of these studies. Arranged according to specialized areas, an overview would be obtained which would benefit many budding information projects, e.g., those ranging from one involving equipping libraries to one designed to find the best way to build up an information service. In both of the cases mentioned above, calculation of the active life-span of scholarly literature should also be mentioned as an important consideration.[9] This factor often causes a limitation of the literature collection in terms of time-span, while the application of Bradford's law permits a limitation of the range by supplying the essential core journals.[10] These two factors may also be deduced by analyses of bibliographies, as discussed in Brookes's paper in this issue.

If question 4 above is extended somewhat further, one can also convincingly distinguish trends in specialized areas. Methodical analysis of the specialized bibliographies may also be applied here.[11] An unpub-

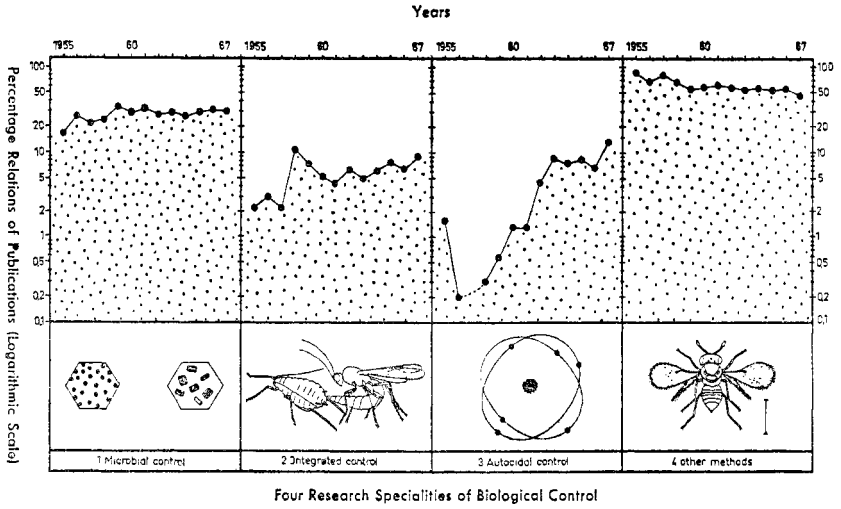

FIG. 1. Percentage Relations of Publications in Four Research Specialties of Biological Control. Data taken from "Bibliography on Biological Control," which appeared in the journal *Entomophaga* (Paris).

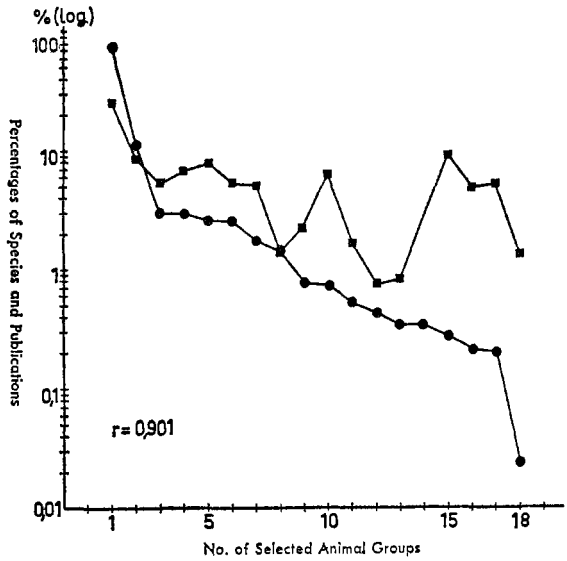

FIG. 2. Relationship between the Number of Species to Publications in Eighteen Animal Groups. Species are signified as circles, publications as squares. Data are taken from *Zoological Record* (London) issues from 1955-1967.

lished example from my own current studies can demonstrate this in the field of biological control of pests (see figure 1). First of all the information flow is represented purely descriptively in one area of research.

In addition to determining how much, where, and what, it is also important to determine the factors which are responsible for course and strength of the information flow. Some very promising formulations in this area have been furnished by de Solla Price.[12] A comparison to this kind of investigation at the international economic level would be the gross national product. Correlations with journals in the "environmental protection" sector are thereby clearly revealed.[13]

In the field of biology the number of species within systematic categories are good comparative parameters for the corresponding share of literature.[14] An example of this is represented in figure 2, which shows the highly significant positive correlation of 0.901 for the numbers of species and the corresponding share of publication for 18 selected animal groups (for details see reference 14).

Introduction

Citation indexes have been a major aid in tracing tendencies and sociological and historical interdependencies in various areas of research.[15] Although it has been found preferable first to investigate the sciences, the *Social Sciences Citation Index,* scheduled to be introduced in 1973, should also provide an improved basis for corresponding studies in the arts and the humanities.

Even the few examples given in this short summary should show that the analysis of bibliographies is a method which, from the viewpoint of information studies, deserves urgent priority in new projects. Analyses of bibliographies can give valuable insights and overviews not only to the information specialist himself, but also to the specialized researcher. This issue of *Library Trends* is intended to be a preliminary, useful overview of the subject and will possibly give an incentive to further studies in the field of "analyses of bibliographies."

References

1. Malin, Morton V. "The *Science Citation Index*®: A New Concept in Indexing," *Library Trends,* 16:374-87, Jan. 1968.
2. Lancaster, Frederick W. "The Evaluation of Published Indexes and Abstract Journals: Criteria and Possible Procedures," *Medical Library Association Bulletin,* 59:479-94, July 1971.
3. *Ibid.,* p. 484.
4. Simon, H. R. "Zur Analyse Biologischer Bibliographien," *Heidelberger Jahrbücher,* 15:111-33, 1971.
5. Wyatt, H. V. "When Does Information Become Knowledge?" *Nature,* 235: 86-89, Jan. 14, 1972.
6. Kennan, Stella, and Atherton, Pauline. "The Journal Literature of Physics: A Comprehensive Study Based on Physics Abstracts." New York, American Institute of Physics, 1964.
7. Anderson, Sydney, and Van Gelder, Richard G. "The History and Status of the Literature of Mammalogy," *BioScience,* 20:949-57, Sept. 1, 1970.
8. Frick, Barbara F., and Ginski, J. M. "Cardiovascular Serial Literature: Characteristics, Productive Journals, and Abstracting/Indexing Coverage," *Journal of the American Society for Information Science,* 21:338-44, Sept. 1970.
9. Line, Maurice B. "The 'Half-Life' of Periodical Literature: Apparent and Real Obsolescence," *Journal of Documentation,* 26:46-54, March 1970.
10. Goffman, William, and Morris, Thomas G. "Bradford's Law and Library Acquisitions," *Nature,* 226:922-23, June 6, 1970.
11. Rothman, Harry, and Woodhead, Michael. "The Use of Citation Counting to Identify Research Trends," *Journal of Documentation,* 27:287-94, Dec. 1971.
12. de Solla Price, Derek J. "Measuring the Size of Science," *Proceedings of the Israel Academy of Sciences and Humanities,* 4:98-111, 1971.
13. Simon, H. R. "Informationsmöglichkeiten zum Umweltschutz," *Natur und Landschaft,* 47:252-55, Sept. 1972.

14. ———. "Animal Species and Publications," *Journal of the Washington Academy of Sciences.* (In press)

15. Weinstock, Melvin. "Citation Indexes." *In* Allen Kent and Harold Lancour, eds. *Encyclopedia of Library and Information Science.* Vol. 5. New York, Marcel Dekker, 1971, pp. 16-40; and Garfield, E. "Citation Indexing for Studying Science," *Nature,* 227:669-71, Aug. 15, 1970.