BIBLIOMETRIC ANALYSIS OF ISI'S ARTS & HUMANITIES CITATION INDEX

Morton V. Malin and Martha C. Dean
Institute for Scientific Information

The most frequently cited journal articles in the Arts & Humanities Citation Index (A&HCI) are analyzed in terms of their disciplinary classification. Four years (1976-79) of the A&HCI data base yielded 144 journal articles which were cited 10 or more times during the period. These articles are predominantly from the disciplines of Language and Linguistics (31%), Philosophy (23%), History (13%), Religion (8%), and Archeology (6%).

However, most citations in the A&HCI are to books rather than journal articles (96% versus 4% among those items cited 10 or more times). The discipline of Literature (or Literary Criticism) is predominant in the list of highly cited books.

These statistics suggest systematic variation in the resources used by the different disciplines. Various other aspects of the A&HCI data base are explored, and some specific examples are discussed in depth.

BIBLIOMETRIC ANALYSIS OF AMERICAN HISTORY DATA BY FAMULUS

Miranda Lee Pao
Case Western Reserve University

Bibliometrics, the quantitative study of literature, has made a considerable contribution to the management of scientific libraries and to the understanding of the sociology of science. It has joined operations research as effective management tools in making substantial impact on optimization of library services, resources sharing and allocation in library networks. With the common belief that humanistic literatures are different, these analytic techniques have not been applied to any extent. We believe that although sociological factors may impact on the humanities in different ways, patterns of regularity in communication, individual contribution, and literature growth shall be studied. Knowledge of the characteristics of humanistic literatures may enable us to improve information access and to maximize available resources in space, manpower, and funds for books. A pilot study has been attempted on a bibliography of the history of the American Revolution.

The data base is constructed from the Writings on American History: A Subject Bibliography of Articles under the heading "Revolution & Confederation (1763-1789)". To facilitate the analysis, the data management program FAMULUS was used.

Selected Results: 1095 articles were published between 1962-1976 by 790 authors in 224 journals or publishing sources. The average author productivity is 1.39 which is similar to other fields studied. As expected, a few highly prolific authors dominated the field, and their productivity distribution follows the well-known Lotka's Law. The dispersion of publications over journals also adheres to Bradford's distribution, identifying the most productive journal in this field as the William and Mary Quarterly. Thus, scholarship in general assumes certain common characteristics regardless of the discipline. Yet differences exist. There are only 77 (10%) authors who have ever co-authored in a mere 47 (4%) of the total literature. This is a dramatic departure from the intensely collaborative enterprise of science in which almost all scientific writers co-author.

From our experience, FAMULUS is a useful tool in bibliometric analysis in addition to being a good personal documentation system to replace our shoebox of cards. We have also used it in the teaching of indexing and retrieval of documents.

A TEXT-RETRIEVAL SYSTEM USED IN HUMANISTIC ARCHIVE APPLICATIONS

Knut Hofland
Sigbjorn Arhus
Norwegian Computing Centre for the Humanities

NOVA*STATUS is a text-retrieval system which is available at all Norwegian universities and several government institutions, running on computers from different manufacturers.

NOVA*STATUS is a full-text retrieval system originally developed by AERE, Harwell, England, and redeveloped at the Norwegian Computing Centre for the Humanities and other Norwegian institutions. The data is divided into documents and each word (or a truncated part of it) in each document is potentially a key word. The format of the document is free. By use of prefixes it is possible to divide each document into specific fields of information.

A request to the system can consist of a Boolean expression of words and prefixes and relational expressions between prefixed words. The system allows for macros which can be permanently stored which makes, for example, synonym lists possible. The system also contains procedures for off-line sorting and printing of catalogues and for the coding of data for statistical analysis by SPSS.

At the Norwegian Computing Centre for the Humanities, NOVA*STATUS has been, and is still being used in a variety of humanistic archive applications. Several of these have a common data format which consists of 20-30 defined fields of fixed information and one or more fields of

free-text description of e.g., photographs, paintings, archaeological and cultural artifacts, old buildings and documents. The second part of the paper will describe the actual use of the system in these various applications.

SIFT - SEARCHING IN FREE TEXT A TEXT RETRIEVAL SYSTEM

Oystein Reigem
Norwegian Computing Centre for the Humanities

The SIFT project is aimed at developing an advanced text retrieval system possessing the features of high modularity, high portability, possibilities for integration with word processing systems and a flexible user interface. Possible applications for such a system would be found wherever any sizable collection of information requires efficient retrieval. The SIFT system is mainly designed to solve the problems of searching in free, i.e. unstructured, text but extensive functions for dealing with structured information are also offered.

The SIFT project is based on former experience in the use of other retrieval systems, particularly the Norwegian version of the British STATUS system, NOVA*STATUS, a system which has found application in various public agencies and at all Norwegian universities.

The SIFT project was initiated on January 1, 1980, and a prototype version of the system will be implemented on a NORD computer towards the end of 1981. The final product will be made available free of charge.

This presentation will treat the structure, characteristics and applications of the SIFT system.

THE TEXT'S THE THING: CONCORDANCES TO LITERARY TEXTS

Michael Preston
University of Colorado

The history of computer-generated concordances is already one-third of a century long. Thousands of concordances have been generated; many have been published. Most of these are useful, but there are limitations to all of them. In this presentation I discuss a number of variations on concordance-making based on specific projects being carried out at the University of Colorado.

A word-form concordance can be of considerable utility. Particularly for older states of language of which our knowledge is often less than perfect, this "primary" concordance form seems best for initial circulation, but such a concordance is insensitive to variants and ambiguities. It is often as suggestive of what might have been done as it is directly useful.

With the increasing availability of microcomputers and various kinds of remote terminals, it is now possible to remove many of the difficulties of text-editing so that a "secondary" concordance edited toward particular applications can be produced more readily. At the University of Colorado, at which the majority of humanists who use computers wish to make maximum use of the available technology without becoming computing scientists, I have found it practical to suggest a particular synthesis of batch and interactive computing. This involves the use of a retrieval, concordance-generating, and editing system so modular in design that editorial intervention is practical at many points. This editing makes use of device-dependent text editors of sufficient sophistication that the user perceives little of the technical operation beyond requesting his programs and his text; otherwise he has the freedom of a typewriter coupled to the benefits of a screen for displaying modifications to his text as they are made, whether directly by him or by a variety of programmed functions. Stations built around "smart" terminals as well as "dumb" terminals with microcomputer and floppy disks are operational.

Thus it is now more practical to produce second-generation concordances which more nearly reflect the perceived needs of a scholarly community: words may be (manually) disambiguated by meaning and function, contexts may be edited either to omit extraneous material or insert explanatory matter, and words may be clustered by dictionary or thesaurus. The result is concordances of far greater utility in specific areas and more meaningful statistics.

The development of better equipment and new techniques has made it possible to interact more thoroughly with one's text. There is no need for premature data reduction, but rather the encouragement of what I call the "infinite loop of literary scholarship": one works with one's texts to produce results which suggesst work to produce more results which suggest still more work.... The newer technology seems to fit the humanist far better than did the old.

A THESAURUS FOR CANADIAN ICONOGRAPHY

Denis Castonguay
Public Archives of Canada

The Picture Division of The Public Archives of Canada has undertaken the construction of a thesaurus of iconographic terms as part of its preparations for a computerized inventory system. The thesaurus which complements an existing set of