Information Retrieval
Applications to Political Science

KENNETH JANDA

Northwestern University

THE BOBBS-MERRILL COMPANY, INC.

Indianapolis and New York

JAMES A. ROBINSON
The Ohio State University
Consulting Editor

To my mother and father

Preface

The term "information retrieval" is comparatively new in political science. In a general way, it refers to the use of mechanical or electronic devices for solving problems of gathering, storing, indexing, retrieving, and disseminating, information. In contrast to the term "data processing," which is used mainly with reference to manipulation and analysis of *quantitative* data,[1] information retrieval has come to be reserved largely for automatic processing of textual material, although the term is sometimes used for accessing quantitative data collected through surveys, censuses, and other forms of research.[2]

Like the methodology of data processing, techniques of information retrieval promise to assume an important place in the activities and concerns of political science, for they offer the political scientist new methods of handling the vast flood of written materials that constitute both the data and the findings of political research. By providing researchers with desired information upon demand, these techniques can be powerful tools for both the production and consolidation of political knowledge. By dealing with textual information rather than quantitative data, they fulfill a special need for political scientists, who rely heavily on qualitative materials that resist easy translation into quantitative form.

What, exactly, are these information retrieval techniques and *how* can they be applied to political science? These questions deserve more searching, comprehensive answers than this short book presents. A thorough discussion of information retrieval methodology requires a

[1] See Kenneth Janda, *Data Processing: Applications to Political Research* (Evanston: Northwestern University Press, 1965).

[2] See Ithiel de Sola Pool, *et al.,* "Computer Approaches for the Handling of Large Social Science Data Files," a progress report to the National Science Foundation on Grant GS–727; Center for International Studies, M. I. T., January, 1967.

book of its own; and it is too early to predict how the full range of information retrieval techniques can be applied to substantive concerns in political science. Therefore, this book will not attempt exhaustive answers to "what" and "how" questions of information retrieval for political science. Instead, the selections try to present answers that are illustrative of the technology and its applications.

A number of good texts exist on information retrieval in general. These include Allen Kent, *Textbook on Mechanized Information Retrieval* (New York: Wiley, 1962); Joseph Becker and Robert Hayes, *Information Storage and Retrieval* (New York: Wiley, 1963); and Charles P. Bourne, *Methods of Information Handling* (New York: Wiley, 1963). Much of the methodology, however, resides in conference proceedings, many of which have been published. The proceedings of a 1964 conference at the University of Pittsburgh, for example, are published in Allen Kent and O. E. Taulbee (eds.), *Electronic Information Handling* (Washington: Spartan Books, 1965). Of special value are the published conference proceedings of the American Documentation Institute, beginning with 1963 in particular. Noteworthy publications, conferences, and other activities in information retrieval are monitored in the American Documentation Institute's *Annual Review of Information Science and Technology*. Volume I in this series appeared in October, 1966, under the editorship of Carlos A. Cuadra (New York: Wiley).

American Documentation, the quarterly journal of ADI, is an important source for articles on information retrieval methodology. Two invaluable guides to both books and periodical literature on information retrieval are *Documentation Abstracts* (quarterly) and *Computing Reviews* (bi-monthly). In its "Applications" section, *Computing Reviews* regularly prints abstracts under the sub-category of "Social and Behavioral Sciences." Occasionally, an abstract is listed under the sub-sub-category, "Political Science," but a better source for articles of interest to political scientists is *Social Science Information,* a quarterly journal published by the International Social Science Council. The better-known journal, *Behavioral Science,* also publishes occasional articles on information retrieval in political science in its section on "Computer Applications."

The most important and probably most accessible source of articles on information retrieval applications, however, is the special issue of *The American Behavioral Scientist* of June, 1964, edited by Ted Gurr and Hans Panofsky, on "Information Retrieval in the Social Sciences: Prob-

lems, Programs, and Proposals." Of comparable significance, perhaps, in its sequel, the two-part issue of *The American Behavioral Scientist* (January–February, 1967) devoted to "Advances in Information Retrieval in the Social Sciences." These issues contain articles concerned more with the substantive application of information retrieval methodology than with its technical aspects.

The selections in this volume emphasize the applications rather than the technology of information retrieval. Even the three chapters on "Techniques" in Part I reflect the development of techniques in response to specific research needs, and these techniques are reported in the context of their research applications. Chapter 1 discusses the technique of keyword indexing with reference to its use in preparing a cumulative index to the *Midwest Journal of Political Science* and in compiling a bibliography on research methodology; Chapter 2 describes the TRIAL program for coordinate indexing by illustrating its applications to propositionalizing the literature on political participation; and Chapter 3 introduces the MIRACODE technique for retrieving information from 16mm. microfilm by referencing its use in comparative research on political parties.

The five chapters in Part II, "Research Applications," employ variations on the three techniques presented in Part I. Chapter 4 describes in detail the computer-generated *Cumulative Index to the American Political Science Review, Volumes 1–57: 1906–1957*. Chapter 5 reviews the advantages and disadvantages of keyword bibliographies on international and comparative politics in comparison with document retrieval services that can search texts for logical combinations of keywords. Chapter 6 treats in greater depth a document retrieval service, called SDI, which was briefly introduced in Chapter 5. Chapter 7 returns to keyword indexing and outlines its applications in identifying and retrieving legislative votes for roll call analysis. Chapter 8 describes how several information retrieval techniques can be used in conjunction with one another in a large-scale research project on comparative political parties.

The three chapters in Part III, "Administrative and Decision-Making Applications," are more speculative than the others. Chapter 9 suggests how a computer program might be used in automating future editions of the *Biographical Directory* of the American Political Science Association. Chapter 10 makes some recommendations concerning the use of information retrieval techniques in a state legislature, and Chapter 11 proposes a more ambitious information system for the United States

Congress. Finally, the Index to the book itself has been prepared on the computer with the program discussed in Chapter 1.

Most of the material in these chapters has been previously published in journals and books, as the source notes to the chapter headings indicate. Many of these publications, however, lie outside the mainstream of political science literature. Therefore, the purpose in assembling this material in a single volume is to publicize the possibilities of information retrieval in political science. As in all collections of articles that have been prepared for publication elsewhere, there is bound to be some overlap, irrelevancy, and obsolescence in the assembled material. Through revision and abridgment, I have tried to eliminate as much as possible of the obsolete and irrelevant material. At the cost of some repetition, overlapping discussions were retained in different chapters if they seem to facilitate understanding.

Because this book is a collection of articles describing one person's applications of information retrieval techniques, it omits some important applications in political research. The most obvious omission is the neglect of computer programs for identifying and retrieving interview questions and corresponding data from large numbers of sample surveys. In this connection, the work of Erwin Scheuch and Philip Stone, Ralph Bisco, and Hans Klingeman is especially noteworthy. Articles describing their approaches to the problem of retrieving quantitative information from data archives can be found in *Social Science Information,* February and April–June issues, 1967.

Although not strictly classified as information retrieval, the "General Inquirer" program for content analysis deserves special mention as a computer technique of information processing that is not represented in this book. Use of this program for analyzing the content of political documents is perhaps best represented in the work of Ole R. Holsti, discussed in part in "An Adaptation of the 'General Inquirer' for the Systematic Analysis of Political Documents," *Behavioral Science,* 9 (October, 1964), 382–388. Holsti's work and related applications are reported in Philip J. Stone, *et al., The General Inquirer: A Computer Approach to Content Analysis* (Cambridge: M.I.T. Press, 1966), which is the major publication growing out of several years' experience with the computer system developed by Stone and his associates.

The case for information retrieval in political science becomes even stronger when it is realized that the techniques and applications described here do not exhaust the universe of existing techniques and applications. Moreover, the present universe must be only a small subset

of future possibilities. The whole range of present and future developments in information retrieval is the concern of a number of special interest groups and committees, such as the Council of Social Science Data Archive's Information Retrieval Committee, chaired by political scientist Carl Beck. While the average political scientist cannot be expected to keep up with all the developments in information retrieval, he should keep a watchful eye toward their successful applications and should be prepared to exploit these applications to benefit his own scholarly activities.

Evanston, 1967   Kenneth Janda