

CONTINUITY AND INNOVATION IN REFERENCE RETRIEVAL IN THE SOCIAL SCIENCES:

Illustrations From The Universal Reference System

ALFRED DE GRAZIA

The computerized Universal Reference System, now independently organized, had its beginnings in the ABS bibliographic reference tools, such as the ABS Guide to Recent Publications in the Social and Behavioral Sciences and "New Studies," which is a regular feature of the ABS. Alfred de Grazia, founder and former editor of the ABS, is Professor of Government at New York University. This article, reviewing the URS and automated reference services in general, is excerpted from his address at the American Library Association Convention in July, 1966.

■ The Universal Reference System (URS) is aimed at achieving the desires of bibliographers, librarians, and scholars.¹ It is a computerized documentation and information retrieval system employing citations of materials above a modest level of quality, appearing in all of the social sciences, annotated, and indexed by author; it includes a set of standard descriptors that are derived from a master system of topics and methodologies and from the unique facets of the works being screened.

The first volume of the URS series on "Political Science, Government & Public Policy" is *International Affairs*, published in 1965.² *International Affairs* is a 1,200 page computerized index to some 3,030 selected books, articles, and pamphlets on foreign politics and international relations. Its value derives in part from the depth of its indexing.³ Whereas most keyword indexes to bibliographies rely solely on titles (perhaps augmented with additional keywords), the URS input is not only annotated but is also tagged with an average of twenty "Standard" and "Unique" descriptors per item. Moreover, from two to four of these descriptors are identified as "Critical" descriptors and are given special treatment in indexing. To illustrate the nature of the punchcard input to URS, some of the 3,030 entries from its "Catalogue" of contents are reproduced in Figure 1.

In order to save space in the computer printout, most of the URS descriptors have been truncated in code words — prompting the volume to be called a "CODEX" instead of an "index." The program for processing the URS entries and producing the keyword index was written for the IBM 1401/1410 computers. Essentially, the program is a type of keyword indexing,⁴ using the descriptors as keywords. Sample output from the URS *International Affairs* CODEX is given in Figure 2.

The section of the printout chosen for reproduction in Figure 2 contains the items tagged with COMPUT/IR ("mechanical and electronic developments for information retrieval") as a descriptor. Of the twelve entries tagged with COMPUT/IR, two have the code word as a "Critical" descriptor (listed in the second column). Assuming the CODEX user is interested in only these two items, he would note their identification numbers — AO182 and AO610 — and obtain complete citations and accompanying annotations from the CODEX Catalogue (Figure 1).

AO181-A0193

EXAMINATION OF THREE ECONOMIC SYSTEMS AS ALTERNATIVE AND COMPETITIVE MODELS FOR THE ECONOMIC DEVELOPMENT OF UNDERDEVELOPED COUNTRIES, PARTICULARLY OF THOSE IN ASIA. SUGGESTS THAT THE INDIAN EXAMPLE OF HIGH PRIORITY TO WELFARE GOALS AND LIMITATIONS ON MARKET FORCES BY PARTIAL PLANNING AND SOME DIRECT CONTROLS MIGHT BE FOLLOWED.

CHINJAP S/ASIA EC/UNDEV ECO/DEV MARKET WEALTH...ECO DOC/ANAL.

AO182 NORTH R.C. ET AL.

CONTENT ANALYSIS* A HANDBOOK WITH APPLICATIONS FOR THE STUDY OF INTERNATIONAL CRISIS.

EVANSTON* NORTHWESTERN U. PR., 1963. 182 PP., \$2.95.

A VALUABLE INTRODUCTION TO A USEFUL RESEARCH TECHNIQUE. INCLUDES CONCRETE ILLUSTRATIONS PLUS GUIDES 'FOR DECIDING WHETHER, WHEN, AND WHAT FORM OF CONTENT ANALYSIS SHOULD BE USED.' THE EXAMPLES COME FROM RESEARCH ON THE ORIGINS OF WORLD WAR I AND ON CONTEMPORARY SINO-SOVIET RELATIONS. SPECIAL FORMS DISCUSSED INCLUDE* THE CONVENTIONAL FREQUENCY COUNT AND QUALITATIVE IDENTIFICATIONS, O-SCRIPTS, PAIR COMPARISONS, AND EVALUATIVE ASSERTION ANALYSIS.

CHINJAP COM EUR+WW1 MOD/EUR USSR INT/ORG ROUTINE TEC/DEV COMMUN PERCEPT RIGID/FLEX ALL/VALS COERCE DOMIN EDU/PROP.... POL QUANT TESTS CON/ANAL COMPUT/IR SIMUL GEN/LAWS GEN/METH.

AO183 LERCHE C.O. JR., SAID A.A.

CONCEPTS OF INTERNATIONAL POLITICS.

ENGLEWOOD CLIFFS* PRENTICE-HALL, 1963. 314 PP., \$4.95.

SYSTEMATIC DEVELOPMENT OF THE CONCEPTS OF THE STATE AND FOREIGN POLICY, THE NOTION OF POLITICS AS THE CRUCIAL FORM OF INTERSTATE RELATIONSHIPS, AND SUBSTANTIVE PROBLEMS OF INTERNATIONAL POLITICS SUCH AS WAR, IDEOLOGY, TECHNOLOGY, TRADE, AND COMMUNICATION.

WOR+45 WOR-45 INT/ORG INT/TRADE DELIB/6P EX/STRUC LAW ROUTINE TEC/DEV ATIT ORO/FREE PWR RESPECT COERCE ECO/TAC LEGIT WAR...POL DOC/ANAL STERTYP GEN/LAWS.

Figure 1. CATALOG OF ITEMS FROM *INTERNATIONAL AFFAIRS*, VOLUME I OF THE *UNIVERSAL REFERENCE SYSTEM*, PAGE 14

International Affairs is URS's first published volume. It is intended that URS become not only a publisher of as many major and minor disciplinary and problem-oriented bibliographies as there may be a need for in the social sciences, but also make available an answering service to individual scholars, librarians, and other agencies, by largely automatic means.

Many claims are made for the system as initiated and planned, but it would be immaterial as well as presumptuous to go into them here. Rather I should like to report and generalize some of the experience that we have had in building a new operation for the time-honored world of bibliographic research and in introducing innovations.

CONTINUITY IN REFERENCE RETRIEVAL

It is said that every invention has more elements of the old than of the new and I hasten to affirm this proposition with regard to the URS. I do so in order not only to credit the past,

COMMUNIST COUNTRIES (EXCEPT CHINA)....SEE COM			
COMPANY, LARGE OR MIDDLE-SIZED (50 EMPLOYEES OR MORE).... SEE LG/CO			
COMPANY, SMALL (50 EMPLOYEES OR LESS)....SEE SML/CO			
COMPETENCE, PRACTICAL (AS GOAL)....SEE SKILL			
COMPUT/IR....MECHANICAL AND ELECTRONIC DEVELOPMENTS FOR INFORMATION RETRIEVAL.			
COMPUT/IR B*51	ATTIT SAMP	WOR+45 WOR-45 COM/IND R+D SOCIETY ACT/RES CHOICE CREATE DELIB/GP PERSON KNOWL....STAT DEEP/INT STAND/INT SAMP/SIZ CHARTS STERTYP	A1322
COMPUT/IR B*52	PERSON PSY	FUT WOR+45 CULTURE COMMUN PERCEPT.... CONT/OBS REC/INT PERS/TEST SAMP LAB/EXP GEN/METH	A1023
COMPUT/IR B*58	TEC/DEV SOC CAP/ISM	FUT USA+45 USA-45 ECO/DEV LABOR SOCIETY ACT/RES PLAN ROUTINE WORK ALL/VALS ECO/TAC....ECO STAT TREND CHARTS SOC/EXP	A1532
COMPUT/IR B*59	COM STRUCT USSR	MCD/EUR ECO/DEV SOCIETY PLAN ATTIT SOC/ISM COERCE COLD/WAR ECO/TAC EDU/PROP....ART/METH ECO HIST SOC CONCPT STAT DOC/ANAL GEN/LAWS GEN/METH ANTHOL	A0805
COMPUT/IR B*60	INT/ORG JURID SPACE INT/LAW	FUT WOR+45 ADJUD CONSTN CONSULT PLAN STRUCT TEC/DEV RIGID/FLEX KNOWL ORD/FREE PWR....TECHNIC GEOG MGT NEW/IDEA TREND DOC/ANAL	A0457
COMPUT/IR B*60	MARKET ADJUD EEC INT/LAW	EUR+WW1 FUT ECO/DEV INT/TRADE NAT/G SOCIETY VOL/ASSN ADMIN CONSULT PLAN STRUCT TEC/DEV ATTIT RIGID/FLEX PWR ECO/TAC....ECO MGT NEW/IDEA STAT TREND SIMUL	A0591
COMPUT/IR B*63	WOR+45 POL/PAR	COM/IND EC/UNDEV ECO/DEV NAT/G SECT CHOICE CONSTN CT/SYS EK/STRUC FORCES LEGIS TOP/EX STRATA PWR WEALTH....POL STAT CON/ANAL CHARTS SIMUL	A0147
COMPUT/IR B*63	COMPUT/IR USSR	CHINJAP COM EUR+WW1 MOD/EUR INT/ORG ROUTINE TEC/DEV COMMUN PERCEPT RIGID/FLEX ALL/VALS COERCE DOMIN EDU/PROP....POL QUANT TESTS CON/ANAL SIMUL GEN/LAWS GEN/METH	A0182
COMPUT/IR S*59	COMPUT/IR DECISN DIPLOM	FUT WOR+45 NAT/G R+D ACT/RES CHOICE CREATE ATTIT PWR....POLICY POL CONCPT STYLE KNO/TEST TREND HYPO/EXP GEN/METH	A0610
COMPUT/IR S*63	POL DECISN	FUT UNIV CONSULT PLAN SKILL COERCE DETER DOMIN EDU/PROP WAR....QUANT HYPO/EXP	A1219
COMPUT/IR S*63	PLAN TEC/DEV USSR	COM FUT USA+45 DIST/IND ECO/DEV MARKET RIGID/FLEX CAP/ISM SOC/ISM WEALTH ECO/TAC....MATH STAT TREND DOC/ANAL	A1767
COMPUT/IR S*64	FORCES SIMUL WAR	USA+45 NAT/G ACT/RES TEC/DEV TOP/EX KNOWL ORD/FREE PWR COERCE....PHIL/SCI POL MATH TIME/SEQ CHARTS LAB/EXP DECISN	A0285
CON/ANAL....QUANTITATIVE CONTENT ANALYSIS* OBJECTIVE, SYSTEMATIC AND QUANTITATIVE DESCRIPTION OF THE MANIFEST CONTENT OF COMMUNICATION.			

Figure 2. INDEX OF DOCUMENTS FROM *INTERNATIONAL AFFAIRS*, VOLUME I OF THE *UNIVERSAL REFERENCE SYSTEM*, PAGE 395

but to apologize for the limitations of the invention. We make a point of separating books and articles in our Index because that is the way that the bibliographers have worked and that is the way in which documents are published; our first products are printed and bound in conventional form because that is the way that most bibliographies are published. We use a considerable number of the descriptors in our system because these are the locating words that researchers and librarians have ordinarily employed to index documents. We limit our capacities and our designs because the market for bibliographic materials is not yet willing to pay for their cost of development and production.

Yet even if we wished to move out as fast and far as possible we should discover that the new invention is not giving people what they want. As soon as a new model of information

machine is offered, some people who have been satisfied with a horse and buggy want an air-conditioned amphibious vertical takeoff Cadillac.

The idea of unlimited capital spending for the development of reference retrieval systems has inspired the imagination of many observers of the developing scene, and one of the problems an innovator must contend with is the demand for complicated and high-cost services on a low payment plan. For example, the type face of the Universal Reference System thus far has been a clear but conventional computer print, all capitals and not unlike a typewriter. There is already available a machine that would use lower as well as upper case letters, but the reprogramming and reorganization that would be necessary is too much for the cost structure to bear at the moment. The several thousand dollars of extra cost involved would mean nothing under a large scale government subsidy because it would not have to be passed along to the purchaser.

Other criticisms levelled at the URS and other systems can be met only by much more substantial costs. Thus, to introduce automatic annotating and indexing of documents would require as much or more perhaps than the development of the URS to date (about \$50,000, that is). I should say that there is a real question here as to the wisdom of using the computer to read and abstract documents, granted the brilliance and utility in other respects of the work of Philip Stone, who has produced the General Inquirer System,⁵ and other scientists. In the immediate future, at least, automatic indexing and annotating are not flexible and subtle enough for our purposes, and the make ready time for the material to be processed by machine is in itself too costly.

The continuous process of relating innovations to the habit structures and expectations of the using clientele does not always present such difficult problems, even in the immediate sense. A couple of thousands of dollars and the inescapable requirement of ingenious programmers and designers should enable us to organize the second CODEX of the URS somewhat differently than the first without losing the continuity and use of the materials already put into the system in the process of producing the first volume.

It does not take much to organize the catalogue by the author's last name instead of by number, as was the case in the first CODEX. This will eliminate the special need for an author's index and permit one to look up an author's name to see how he is represented in the bibliography. Nor is there much difficulty involved in carrying the author and title of the work along with every single index entry governing a document; here the problem is one of the utility of printing more descriptors as opposed to printing the title.

Here is how the index of the URS Codex II on *Legislative Processes, Representation, and Decision-Making* will probably be set up:

INTELLIGENTSIA

B67/A4001

DILLON, B. CAUSES OF PROTEST IN OUR TIMES
US+45 VOL/ASSN FACE/GP PERF/ART ...INT/ORG ATTIT
PEACE EC/DEV EC/UNDEV SUPREGO....
MYSTIC POL ART/METH HUM BIOG

A typical Codex of a field of social science (such as *International Affairs*) will contain about 1,000 pages, 1,500 to 4,000 annotated document entries, perhaps 45,000 index entries of the above type, and will rotate on a biennial schedule. Codex II on *Legislative Processes* is scheduled for publication in the summer of 1967. An updated version of *International Affairs* will also be published later in the new format. Our plans now are to issue eight other "sub-codex" volumes, each with about 500 items collected from 1966 and 1967 references. In the winter of 1967-1968, we intend to publish full-length Codexes on each of the same eight topics, including public

opinion, law, economic regulation, etc. Our experiences in publishing these additional volumes should yield continual improvements in URS products.

Such changes from volume to volume constitute action research with a vengeance; each new publication can be compared with its predecessors. Dictionary cross-referrals can be increased; subcategories can be introduced into the master classification system; and hundreds of proper names can be introduced into the index. One of the serious but corrigible defects of the first CODEX was the failure to devise an inexpensive method of carrying forward into the final printout the references to major personalities and places treated in the documents being surveyed and reported.

In general then the problem of variety in the format of the catalogue and index is a matter for continuous adjustment and all that we need is patience with the current product and research on innovations. Finally, there may arise a consensus on what the best format should be, or better still, the ability to make available on order several format possibilities, any one of which is capable of emerging from the computer printer following a simple instruction.

In this process of experimentation and living with the product, some complaint will simply disappear as it becomes obvious that reference retrieval is not going to bring a paradise in bibliographic research. Thus, the very richness of items in the URS Index has been the target of criticism among some observers who have failed to realize that the ordinary bibliography is so scantily indexed that any given category of index entry simply does not reflect all the works in the catalogue that deal significantly with the index term; this is the proverbial embarrassment of riches. The way out of this is greater qualitative discrimination among items and finer theoretical discriminations among subconcepts in social science so that words have more precise meanings at the same time that words of the same meaning are coalesced. This process of experimentation and adjustment can only lead to the improvement of the ways of thinking and method in social science as a whole, as well as in bibliographical work.

One of the major objections to the Universal Reference System has been that there is no large need for it. The *Public Affairs Information Service*, the *Book Review Digest*, the *New York Times*, and several special bibliographies operate in the field. There is no lack of material to which students can be referred. But the computerized reference services can do useful things, even by the old ideals. Once the library is reorganized with some initial capital investment, it can provide the old services more completely and with less work and expense. At least a million students might each year find their work of preparing lists of titles considerably expedited.

The few creative scholars who do most of the consequential reading and writing will be most helped, however, for they will be able to call upon an instant service that will be many times as powerful as the old way of bibliographic research, exactly analogous to the powerful Cadillac that at a touch of a button puts several hundred horses to straining at the harness.

It is to the advantage then of the best scholars and best librarians to profit from the irrational excesses inherent in any new movement for technological change to acquire equipment which, if they tried to purchase it for themselves, would involve them in long and useless expositions before the budgetary authorities of universities, institutes, and agencies. The computerized bibliographic system is a soul-appeasing therapy for the academic mass and a necessity for the serious scholar.

INNOVATION IN REFERENCE RETRIEVAL

Much work on the subtlety of expression in annotating and indexing remains for the future. When the coder of a docu-

ment is asked to summarize it and then to index it according to the master formula by ticking off the relevant descriptors, he acquires momentarily a concrete sense of the work. But he can only lend part of his appreciation of the document to the system for the final combining, permuting and printing. The syntax, and therefore much of the meaning, is lost in the successive transformations of the material thought. J. C. Gardin and others have worked on the problem,⁶ and I have given it some attention with a view towards lending enriched and precise meaning to the annotations, index, and groups of items.

We found from examining our coders' annotations that they used the same verbs over and over again. A couple of dozen verbs seemed to suffice for half the relationships that they needed for writing a seventy-five word annotation. These would be words like "describes" and "compares" and "analyzes." One step toward a more subtle rendering of the index would be to code such connecting words and reproduce them mechanically between the descriptor topics and methods. One could also experiment with the design of a study space to which the coder could affix descriptors in proper topological relationship to one another, and have the machines reproduce each work's annotation and indexing as a kind of chart.

Our present inclination, however, especially if we are to preserve continuity with the existing programming, is to move in the direction of grouping works by shared combinations of indicators or descriptors, printing out those groupings that are likely to be most useful to a given bibliographic search. Thus, the machines would be programmed to shuffle through their tapes or discs to produce lists of all articles on nationalism in underdeveloped countries before World War II that deal with political parties and leadership and that were the product of field research by sociologists.

The major reason why we have not gone this far already is because the subsequent printouts in a large subject area such as the international area would be so voluminous as to make a single-volume printing impossible. Hence simpler groupings have been resorted to, leaving some of the work of combining to the naked eye of the reader of the CODEX. The reader does the task by checking each item under one of the desired descriptor headings, such as *field research*, goes directly to the articles written before World War II and scans the items in the group for all the desired descriptors. This can be done at the rate of about one hundred a minute, so the loss of time is not excessive. Then the reader or his typist extracts the items containing the combination by resort to the catalogue and prepares his own bibliography on the detailed specific subject of concern. These latter tasks could be taken care of by full automation if there were sufficient demand to create a multitude of lists at a low price.

The question whether to provide bound volumes or lists is not so much one of technique as it is of costs. The final form in which the product can be made available to the user is a decision to be arrived at after many considerations of who the users might be, how many of them there are, how much the tailored product will cost, whether local co-operating units can be set up, whether a subscription system is possible, and so forth.

We are beginning the URS with bound books because that is the conventional way and requires the least break with tradition and habit. The problem of updating is serious only in the cost sense; bound volumes supplementing the original volume cost almost as much as publishing the whole bibliography in a second edition with deletions and additions, both accomplished by machine. The only expense of republishing the whole with amendments will be the cost of the total printing and binding job of, say, a 1,200-page book as compared with a 400-page book. (Or, let us guess, between \$35 and \$30.)

As soon as conditions permit, a subscription and special order service will be established by the URS and presumably by any other system in like circumstances. This would enable those users who are able to afford more sophisticated and complete services to place orders for special listings as referred to earlier. A professor may order for his classes special bibliographies of up-to-date literature, divided more or less according to his outline of lectures or conferences, the URS providing him with a list of works perfectly organized to correspond to each of his major subject headings.

As I have said, we possess this capacity as well as many others in the system at the present time. Only the very slow turnover of working capital delays the fulfilment of such capacities. And I should add that it should be possible for any other group to achieve the same position within a year or two with a hundred thousand dollars, in view of the programs, equipment, experience, and skills that have become available.

Not too far away lies the prospect of console hookups to reference retrieval centers from libraries and other users. I am not so sure of the economics of this setup but I am of the opinion that should a bonanza in the way of a government subsidy be discovered, so that the capitalization problem of research and development might be ignored, the URS or another similar group might combine with groups of advanced technical capacity in the field of information retrieval. They could together establish very quickly a national system of reference retrieval that would let a central computer receive instructions or requests from subscribers anywhere and fill the request by listings on a telescreen, and when corrected, by printing out the desired information in the office of the subscriber. Harold Borko of Systems Development Corporation has been rehearsing these possibilities recently; others may also be on to this kind of system.⁷

The awesome efficiency and effectiveness of such systems as I have envisaged here stand in sharp contrast to some of my cynical remarks about the uses of bibliographies. I would not retract the remarks; I believe that we should not be abashed by our creations. We know, after all, how enormous and complicated is the system that is used to get millionaires to the Virgin Islands for a weekend of restless play; we know how many millions of dollars go annually into the purchase of a few paintings and a few rare books of which copies are readily available; we know what a crazy-quilt of agencies and money

it takes to cure a case of poverty. With these and many other examples in mind, it would be altogether too self-abasing to deny the investment of money and energies required to build some libraries that correspond to and convey an image of modernity and of the future of intellectual work. We shall not expire if all the things we wish libraries or scholars to be do not come about. But we shall be happy if they *do* happen. And it is fun to try to make them over in some new vision. And there is no doubt but that more useful studies will be facilitated than are presently possible.

REFERENCES

1. Alfred de Grazia, "The Universal Reference System," *American Behavioral Scientist*, VIII (April, 1965), 3-14.
2. *International Affairs*, Volume I of the Political Science, Government & Public Policy Series of the *Universal Reference System* (Box 262, Cooper Station, New York, N.Y. 10003: Universal Reference System, 1965), 1205 pp.
3. This description of the *International Affairs* index is adapted from Kenneth Janda, "Information Retrieval: Applications to Bibliographies on International and Comparative Politics," a paper presented at the Computers and the Policy Making Community Institute, Lawrence Radiation Laboratory, April 4, 1966.
4. See Kenneth Janda, "Keyword Indexing for the Behavioral Sciences," *American Behavioral Scientist*, 7 (June, 1964), 55-58.
5. Philip J. Stone, et al., *The General Inquirer: A Computer Approach to Content Analysis* (Cambridge, Massachusetts: M.I.T. Press, 1966). Also see Erwin K. Scheuch and Philip J. Stone, "The General Inquirer Approach to an International Retrieval System for Survey Archives," *American Behavioral Scientist*, 7 (June, 1964), 23-28.
6. Jean-Claude Gardin, "A European Research Program in Document Retrieval," *American Behavioral Scientist*, 7 (June, 1964), 12-16. Also see Jean-Claude Gardin, *Syntol* (New Brunswick, New Jersey: Graduate School of Library Service, Rutgers University, 1965).
7. Cf. H. Borko, *The Conceptual Foundations of Information Systems* (Santa Monica, California: Systems Development Corporation, May 6, 1965) 37 pp. and Noah S. Prywes, "Browsing in an Automated Library through Remote Access," in M. A. Sass and W. D. Wilkinson, eds., *Computer Augmentation of Human Reasoning* (Washington, D.C.: Spartan, 1965), pp. 105-130. Also see Harold Borko and Lauren B. Doyle, "The Changing Horizon of Information Storage and Retrieval," *American Behavioral Scientist*, 7 (June, 1964), 3-8.