

IV

Continuity and Innovation in Reference Retrieval
in the Social and Behavioral Sciences

(with illustrations from the Universal Reference System)

by Alfred de Grazia
New York University

Reference Retrieval is a synonym for bibliographical research. (There is no use to pretend too much.) Every librarian and scholar believes that he or she is God's gift to bibliographic research in his own field, so that we must make it plain that what is said here is by way of recapitulation and reordering and not of novelty.

I have reformulated the eternal process of reference retrieval as follows:

How do we discover who says who
transacts what with whom, where
and when, why, and how does he know so?

1. How do we discover --- What is the technique, who is capable of employing the technique, what resources are needed, how much needs to be discovered, what is the cost of discovery, how efficient must it be? Actually the total framework of this paper is wrapped around the answer to this question. To a relatively lesser degree, we are delving into the remaining parts of the eternal formula, viz.

(June 1966)
delivered before the American
Library Association Convention

2. Who says --- This queries the citation, the name of the author and his distinguishing marks, such as time and place of publication, publisher, number of pages, format of the work and whatever other necessities and luxuries of identification such as his birthdate, his affiliation, the full names of everyone in a symposium and so forth that should and can be accommodated to the medium of the bibliographic output. We remark here how complicated the set of decisions to be made if we are to concern ourselves with bibliography on a large scale. The problem is simplified on the one hand nowadays (e.g., we can add Library of Congress accession numbers) but complicated by the proliferation of publishing sources (how do we standardize and address all the back room mimeograph machines that are turning out products of value?).

3. Who --- This next 'who' is the primary actor of the document. We assume that there always is one and, where we must, we force someone or something onto our procrustean bed. This 'who' may be a corporation as well as an illustrious personage; it may be "the international balance of payments"; it may be God, it may be peace. I have endeavored to classify all the major actors and they may be observed in the topical sections of the classification system that you hold. ¹

The actor is an elusive creature, however, always somewhat abstracted. Is it the raw cabbage that prevents scurvy or the vitamin in the cabbage; it is the vitamin, we say, of course, but for a while cabbages were used on ships at sea and vitamins were still unknown. So it goes with all actors; they are variously categorized, they take on numerous roles; a proper description of the 'who' would give the actor as many descriptive names as he has traits or behaviors that are essential for the performance of whatever action we are trying to distinguish for the purpose of letting someone else know what is going on in the document being described.

Often this listing of traits can be very long indeed. For example, a book on state and local government in America may contain thousands of specific actors -- units of government by species and proper name, organizations, institutions, parties, functions and so on -- until properly delineated, the 'who' will give us a rich list of descriptors equivalent to all the terms of the table of contents, the index of the work and such special descriptors as appear only in our own descriptive system for actors. This is not impossible to provide in an index to a bibliography, but wants, necessity, costs, and bulk considerations intervene to limit drastically the list of such 'who' descriptors in any bibliography. (In passing, I may say that the

future may see various classes of detail mechanized for retrieval beginning with the barest description of the actors ranging through elaborate listings of traits until the document itself is produced.)

4. The next part of the eternal reference formula says "transacts what with whom..."

Let us take up the 'whom' first because it is easier to dispose of. A number of social scientists, myself included, find the concept of transaction a useful and universal one. Each and every social event is a communication between two actors. So, find an actor, and you find someone acted upon; and he himself will be acted upon: such is transactionism. We need not philosophize further here. The point is that a common need in seeking out a document is to discover the predicate of the action.

Now, logically, if our list of types of actors is complete, our list of those acted upon is complete. They are one and the same. The 'who' is the 'whom.' Both can be found in the same list. Voters are interviewed, for example, to discover their opinions about politicians. Voters and politicians are the subjects and objects. (Reverting for a moment, I would recall that each of the two groups has other attributes as actor, which someone may be wishing to discover -- for instance, the public, the majority,

working class people, the educated person, etc., on the one side, and legislators, governors, town councils, professional politicians, etc., on the other side.

It is plain that between the 'who' and the 'whom' there is a transaction that needs to be described for many reasons in bibliographic research. What are the actors up to, what are they doing? Are they thinking, playing, administering, judging, consulting, loving, killing, or simply relating in a statistical sense as when bond prices are being related to stock prices, or what? I am more dissatisfied with my classification system in this respect than I am in general because I intended originally to avoid the verbs connecting the actors in their operations, save in a few key situations. The instruction I gave to myself and others here was to use subjects or actors to describe the action also. For instance, if politicians were bribing voters, I had a rather natural rubric of an active sort called "economic tactics" which means the employment of economic measures or tactics for the purpose of affecting the financial or goods position of the subject. (This was an extremely general action, which is, however, another problem.) But the same action sequence is supposed to be described under the choice category, which, as one can observe in the classification plan, deals with "choice, elections, modes of deciding alternatives or selecting officers and leaders for structural positions." The situation isn't too bad, you see;

any verb can be converted into a noun, and then you would have 'who transacts what with whom' listed minimally under three nominative topical categories.

(I would note again in passing that we are skirting the edge of a garden that we should come back to -- the highly important field where topics in an index are associated as they are associated in "real life". That is, the terms voters, politicians, economic tactics and elections, in the example given, do not really tell us which direction the action is taking in the work under consideration. Even if the term bribery were added as a unique descriptor, we should not know whether the politician is bribing the voter or the voter is bribing the politician. (I might philosophize here that when a politician bribes a voter the voter is also bribing the politician in that he is inducing him to stay away from larger idealistic issues... but I had better keep my role as political scientist out of my function as librarian.)

The search for a work is often a search for a process. These processes have to do with educating, persuading, coercing, dominating, or manipulating someone or some institution. Furthermore these processes are often independent of the actors in the sense that a search for an actor will not produce the actor operating by the special means that one is interested in. Thus a work on the radio industry will not necessarily be suggested when one is seeking works on popular education or on monopoly but if it is categorized by the means the radio

industry employs it will be caught under the categories of manipulative tactics used in the index. The same is true of certain psychological processes that are obviously independent of the actors, such as perception, displacement, drive and the like. To summarize here, then, the URS indexes the transacting process not as verbs but as nouns. This has a major disadvantage in that actors and acting are indistinguishable and one cannot reconstruct the original full sense of the process from the Index alone but has to go to the annotation, where the essential bibliographic query can be answered in prose.

5. Where and when --- The human sciences are locked into a time and space frame that should be reported as an essential part of the reference retrieval formula. The number of proper names for the peoples, nations, cities, battles, and events of history is large, of course. They tend to be grouped in the interests of scholars. A specialist on Ecuador and one on Argentina will have a number of common sources, whatever the differences between the two countries. And there is also a conjunction of time and cultures and nations and events. If one were to reduce the total historical record to a highly abstract time-space scheme, he might end with a set of categories somewhat like that which I have proposed and introduced into the Universal Reference System. There, for example, the communist nations (except China) are given one of the eighteen categories that embrace all of history and include even one for the future and one for universal studies, that is, studies whose propositions are believed to be independent of time-space reference in the

broad sense (for example, the principles of administration or aggressive behavior in foreign affairs).

The advantage in this scheme is definitely afforded the broad student or scholar, not the specialist, and we very quickly determined to introduce unique descriptors to designate particular locales of events treated importantly in the documents. In the end, of course, these will constitute subcategories of the more enduring and spacious categories of the master classification.

6. "Why?" is the next query that occurs in our formula, that is, why does the actor (or author, if the actor is the author) act the way he does or take the position he does take in describing the action? The social sciences are normative sciences and to describe a document without being alert to the value distributions being talked about or the value position of the author would make the effort less worthwhile. And of course the typical classification for library purposes and retrieval does recognize major journalistic categories of value such as communism or fascism, and major ethical philosophies. But only a conventional kind of analysis is performed for there is little inclination to insist that every work can be queried as to its value position and the values it discusses. For example, if an article in the area called group dynamics deals with power relations and a book on the future of democracy deals with power relations, ought not both to be found by a person researching the nature of democratic power?

I have therefore adapted Harold Lasswell's scheme of values to my classification system.² Documents are analysed according to the goods of life for which the actors are striving or in which the author is interested. If the work deals with income in the United States it is indexed on wealth along with another study that may describe the attitudes of high society to the poor. Again we are faced with having many, many works in the same value category. But that is true to life. And besides there are other facets of the works under classification that enable them to be discriminated rapidly in other respects. And furthermore it is healthy for the social science and all sciences, not only in respect to this example but in respect to all concepts, to continuously make new combinations of things that are conventionally dissimilar but basically the same. This is one of the aspects of the CODEX and URS as a research tool that go beyond bibliographic searching.

7. At this point we move into the final large division of the classification system that answers the final portion of the master question. How does he, the author, know so? --- the methodological question. More and more the sciences of man are becoming methodological and operational in that substance and method are intertwined. Traditional and popular classification systems have been woefully deficient in this regard. They may say in their cataloguing or indexing whether a work is in psychology if it is an obvious treatise on psychology, but in

the URS every work that deals with mental health is indexed in psychology, every work that deals with power is in political science, and so on. If we wished to do so, for example, we could run separate lists from the CODEX that could be entitled "The Psychology of International Politics," "The Sociology of International Politics," and so on.

More and more, students must search out the methods of work, with the subject becoming less outstandingly important; indeed in many cases, and justifiably so if we are educating social researchers and social engineers by the hundreds of thousands, the method is the more important facet for the student, not the subject. If it is market research by sampling techniques that a student is concerned with, he may not be happy with looking through all that has been written about the soy bean market, commodities of particular kinds, the structure of the distribution industry and so on. He would rather find documents that deal directly and heavily with the technique of research.

Therefore, the methodology of the work in question came to be an important part of the search formula for documents in the URS. There are the highly general categories of the field or discipline, such as sociology or history, and the narrow more indicative and denotative categories of the sample survey -- the depth interview, programmed instruction, and so on.

I may comment that I see little scholarly or scientific use for the conventional categories but you can imagine how many people counselled me to keep the conventional discipline terms as one facet of indexing. Someday it must come out, I feel. For in classifying a work, of what ultimate use is it to know that it is written by someone who is called an economist? What ~~are~~ important are his subject and method. This is certainly one of the problems of innovation in reference retrieval, one among many.

Now the problem of organizing methodological facets of a work is somewhat the same as the topical problem. There may be a large number of techniques within a given category of technique. I need not remind you, for instance, how many kinds of psychotherapy there are; yet we lump many of them together under depth interviews: simple, psychanalytic, hypnotic, or with drugs. What happens to the scholar or student who is to do his thesis on hypnosis?

Well, if I may be permitted a little irony, he might go to the Index Medicus which has millions of government dollars to spend upon mechanizing the medical literature. Or if he is a psychologist, he will look up hypnosis in the books that stand on his shelf, or go to the new index of Psychological Abstracts presently in process of publication. Or he can wait until the day that the Universal Reference System gets around to psychology and adds a number of unique descriptors creating new subcategories under Depth Interviewing with a correspondingly rich representation of works in the field of hypnosis.

But at the moment he might use the system for what it is by examining the Codex of International Affairs³ to see whether any works appear to exist in the rather short list of items dealing with depth interviewing that contain material on hypnosis. One of the reasons that he might not succeed is that there have been in fact very few works utilizing this technique in the study of international affairs, and perhaps none at all on hypnosis. However, under the entry of psychology, which he might also work with, he would find some near misses such as Gabriel Tarde's The Laws of Imitation or Lasswell's Psychopathology and Politics.

But a moment ago we were talking of values and perhaps it should be noted that one of the first of the sets of descriptors in the methodological section deals with the values of the author. There was much outcry over this rubric, and I can understand it.

No one will deny that there are some obviously partisan works. No one would object to labelling Lenin's Imperialism a communistic work, nor Adam Smith's Wealth of Nations the work of an Old Liberal. But is Ignazio Silone's School for Dictators the work of a socialist? It is, but who gives the coder the right to say so and how could you expect him to be so educated? He might remember that Silone was once a communist. Would not one go quickly into a McCarthyism of bibliographic labelling? Would the term Grazianism become something my children would have to live down?

I think not. Coders are all too eager to evade judgments and flee into the haven made available to unclassifiable works. Meanwhile, small help though it be and often debatable as a choice, the description of a work by the politics of the author can be of great help. For instance, if you wish to see what some communists have to say about economic development, where would you go? To an expert? Yes, that would work especially if he had the time and patience for you, and a good catalogue of his own or a good memory. But why not simply discover it where it should be discoverable, in the CODEX?

The reason why the ideological affinities of the author should be categorized under methodology of the work are probably obvious. So also for the field or discipline, that we have already referred to. Thereupon I have tried to elaborate the most abstract and logical breakdown of the research process to allow for capturing the method of any work and relating it to others of like class.

Thus, biography is included under the analysis of temporal sequences, and includes not only conventional personal histories but studies in personality development and psychoanalysis. This was done in order to excite students into thinking analogically and analytically about personality development. For many purposes -- more so in the future than in the past of social science -- a scholar who wishes to learn about how politicians become what they are will and should seek out not only studies of famous politicians but the often superior studies

of ward leaders, ordinary citizens, and students written by political scientists, sociologists, anthropologists, clinical psychologists, and psychoanalysts. This is not only good social science. It is good history, for the science of the biography of dead men cannot help but be benefitted by the lives of the living as obtained through intensive depth interviewing and direct observation. To emphasize the matter, a good bibliography can and should be a new, directing instrument of investigation.

We are now beyond the explication of the grand formula for bibliographic discovery and ready to explain the general nature of the mechanisms for carrying it forward and into being, but I would pause here to mark off one important new point. My original vision of a bibliographic classification was in many ways my vision of the general classification of social science. The categories of the methodology section of the classification scheme are, for example, the outline of my course in methodology offered to graduate students at New York University. The classification of time-space-culture groupings is what I would use if tomorrow I was to volunteer a course of study in World Civilization. And so on.

But more pertinent to the present discussion is another vision that I had in mind in developing the classification for reference retrieval. I see every social event in the world as composed of a combination of a limited number of facets which if described would place that social event in relation to every other social event in the world. I believed furthermore

that one could retrieve a description of any event by mentally forming a construction of it. This theoretical construction could be made an input into a sorting machine that could rapidly perform combinations and permutations. The machine would discover in its stored data that particular incident together with whoever had described it, and by logical extension every other incident that had ever been described that shared all the qualities of the given incident. This was my theory and ideal of information and reference retrieval. The common qualities of events from all times and places are to be shaped into receptacles for the placement of all actual events, and any desire for the recapitulation of any specified event is to be satisfied by an electrically swift search for all events containing its qualities.

There are many practical and temporary obstacles to this achievement but the ambition remains the same, and the theory should hold not only for reference retrieval but also for information retrieval itself as soon as or wherever the contents of documents, that is, the data in itself, is substituted completely for the documents or referring data. In more specific terms of reference retrieval, the goal of the URS and, I would hope, other systems to come is to select the several pertinent and salient qualities of an item being sought and to compare the whole data bank, the whole set of items, that is, with it until whatever emerges is sought and what is not sought does not emerge.⁴

II. Continuity and Innovation

The Universal Reference System and other similar theoretical and applied systems are described elsewhere,⁵ so that I shall not go into the mechanics of reference retrieval. Everyone here is acquainted with the traditional forms of searching materials in libraries, and with the format of conventional bibliographies published by ordinary composition and printing methods as books and pamphlets. Everyone here knows that in recent years a variety of devices has been employed to make more convenient, to speed up, and to more richly index bibliographic work.⁶

The Universal Reference System is a set of devices containing numerous experiments aimed at achieving the desires of bibliographers, librarians, and scholars. It is a computerized documentation and information retrieval system employing citations of material 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 system as a whole is described in the flow chart that is in your hands.

It is intended that the URS become not only a publisher of as many major and minor disciplinary and problem 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.

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, but to apologize for the limitations of the invention. They say that the automobile has its motor in front because that was where the horse was. 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 airconditioned amphibious vertical takeoff Cadillac. If we say that little is new, less will be expected.

In many cases where people expect a reference retrieval system to do everything for them, their problem in the first place is not knowing what they want. And the librarian necessarily is not able to formulate what they want. And both unfortunately may turn to the god of the machine and say "Retrieve for me, o lord."

Still objections to change are dealt with rather cavalierly by most innovators. They are prompt to imply a negative moral quality to the defenders of the status quo and to their critics.⁷ I believe that just as opponents to change employ many improper tactics and confuse their mood with their reason, innovators do the same. In discussing resistance to change we should be careful to delineate elements of resistance that come from mental quirks and those that stem from proper rational objections within the framework of the problem.

One of the major objections to the Universal Reference System has been that there is no large need for it. The PAIS, the Book Review Digest, The New York Times, and many special bibliographies preempt the field. There is no lack of material

to refer students to. Now one of the premises of this argument and its rebuttal must have to do with the absolute need for a service. This is a most involved problem and we could not hope to solve it here.

It is not enough to point to the "information explosion," as that phenomenon has been called, and say that there is no way except the computerized way to put this vast and rapidly increasing flood under harness to the student and librarian.

Much of the flood consists of duplicate materials and if one keeps on selecting as he has, he will continue to get the same result. Also, a condensing hierarchy comes into being and one often gets accurate-enough reports through the review literature. Moreover, as soon as it is discovered that the mass of additional literature is not being read, much of it will cease to come into being.

Finally, most material of the recent past has gone largely unread, whatever people believe, (and some new evidence can be cited for this)^s so what possibly can be the bad effects of the new enlarged mass being slightly less read on the average. If 10,000 scholars do not read 100,000 pages, and 20,000 scholars do not read 1,000,000 pages, the result is not five times as bad; it is zero in both cases. Or let me say, if the reasonable complaint arises that somebody must be reading something, that if 10,000 scholars read an average of 1000 pages, that is ten million pages, and each page of 100,000 pages is read 100 times on the average; whereas if 20,000 scholars read an average of 1000 pages, that gives 20,000,00

pages that are read, or each page of the new flood of ten million pages is read ten times on the average. I would say that whether the average published page is read ten times or one hundred times matters very little.

After all, if half the new stuff is duplication of the old writings in one way or another, that raises the average readership of the really new stuff to fifty readers or half of what it is today, and if only a hundredth of the new new stuff is good in comparison with a fiftieth of the old stuff (a justifiable statement since the doubling of scholars does not represent the doubling of the number of creative scholars), then the absolute decrease in the worthwhile pages being read probably comes out to an average little different from the average today. That is, in real terms, the flood of literature in the social sciences could be ten times what it is today with little change in the amount of readership each worthwhile article gets. With all humility I offer you this Grazian Index of Real Readership, which can be used by librarians to justify the refusal to purchase the Universal Reference System.

Actually I am onto the true secret of what is bothering us in this information explosion. And if you accept the Grazian Index of Real Readership you can explain the hysteria that still occurs despite the force of its explanatory power. It is that any universe of scholars is not a group of readers primarily but a sociable group. The members of this sociability group, like any other sociability group, talk about each other.

They not only talk about each other at conventions, in class, and to the larger public; they talk about each other in reviews, books, articles, and footnotes. When they no longer read the same things written by the same people, they fall into an anomic condition that has a true psychopathological aspect. They get confused, they drink too much, their hair grows grey, they become professional "drop-outs", they become listless, and they demand bigger computers to regain heaven. What they want, whether they know it or not, is to restore the scholarly community so that they can once more pass the same names and facts back and forth.

With a little more effort, I believe I might persuade practically everyone that the major reason for concern about the information revolution or explosion is the psychic wellbeing of scholars as human beings. Such a belief would be well-nigh disastrous to the innovations which I am espousing, however. Given not only the limited budget of librarians but also their well known misanthropism, why should they make their clients feel better? Yet in all frankness I must confess that there are other reasons than the explosion of information for computerizing and otherwise reorganizing the bibliographic part of scholarly life.

One is that continuation of the present methods will do the same thing: lists will get longer and people will feel just as bad, and mind you, in their hysteria they will blame the librarians for their acute anomie, and demand computers be set up even if computers cannot help. For it is now

becoming obvious that the charming little scientific communities of the nineteenth century on which the whole intellectual organization of society is based are gone and the existing bibliographic services are gone with them.

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 too will each year find their work of preparing lists of titles considerably expedited. That should be a source of relief to the authorities as well as the students in an age when the whole educational establishment is being shaken by reorganization for the sake of mass higher education.

And getting back to the few creative scholars who do most of the consequential reading and writing, those people will be greatly benefitted, 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.

Inherent in the discussion of change and resistances that I have just presented is a second major type of problem which needs separate treatment. I refer to the struggle of quality against quantity in the data bank of the reference retrieval system. The question is commonly addressed to us in the following form: who is to determine whether an item should be included in the system or excluded? The same information explosion to which everyone refers creates a strain on the reception rooms of the scientific establishment: should we order this or that new periodical or book; must we get everything; how do we tell the good from the bad; to whom can we pass the buck, in ordering for reasons of quality not quantity?

The democratic reply to such questions tends to be to include in a bibliography of a subject everything published about it. Else, the argument goes, we shall discriminate unfairly and possibly bury forever great works that are not immediately recognized. How many backwoods Galileos will never reveal themselves if literature reporting services become selective and their works are passed over by semi-professional laborers or even panels of established scholars?

This democratic conservatism becomes an especially powerful motive if it is associated, as it often is, with a

bent towards socialism. Only the government, the argument goes, can afford the tremendous expense entailed in the development and operation of the great system that would be needed to gather the basic data and retrieve the necessary information from the millions of pieces accumulating in the whole literature of the social sciences.

Whereupon the argument picks up force, using consequence as causation: since the government would be undertaking this work, in collaboration of course with the existing scientific institutions, it is all the more important not to permit discrimination on alleged grounds of quality. It would be wrong of the government to exclude anybody since the government is everybody, and furthermore quod omnes tangit ab omnibus approbetur: if a monopoly is created, everybody ought to be guaranteed admission to its corpus. (And, we might add from our sociological knowledge, once committees are set up to determine the rules for admission on matters of low unit importance but high general prestige, they will not find it politically expedient to permit anyone to say that so-and-so are not up to the quality of the bibliographic service in their work.

This is one of the reasons why I do not regard the National Science Foundation as a wholly good influence in the field of information retrieval. At some point in time, the NSF will be in a position to provide a monopolistic solution to the reference retrieval problem. That solution will probably incorporate an undesirable or even malfunctioning

principle governing the quality-quantity problem. I would rather see many non-governmental efforts begin, most of them fail, and a few survive under competitive conditions. Under a pluralistic set-up, a man has a chance to have his work gathered, stored, and retrieved by one or more of several groups and he does not suffer as much, even if all groups exclude him, as he would if the one powerful group authorized by the great white father in Washington says his work will not pass into the sanctified pages of the official bibliography.

And, almost with diffidence, I would conclude this discussion by saying that, in my experience, even fairly low standards of qualitative discrimination will reduce the number of items in the system to a manageable total. I cannot see much of value escaping the URS as ultimately developed. But I would still be happy to have several systems going simultaneously and reporting out what they think is good material on a given topic, carrying their own set of information characteristic of that material.

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 thousands of 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. The sum would mean little even to a large-scale non-governmental company such as Xerox for it would thereby merely extend by a few months the date when the system would pay a profit.¹⁰

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, 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. (I think it is understood by everyone that one of the advantages of the URS and similar systems is an interchangeable collection of items in the data bank.) Every innovation has to be judged by the extent to which it incurs excessive costs when it makes obsolete the already developed machinery and data.

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 permits 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 title of the work next to 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. 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 print-out 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 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 sub-concepts 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.

Much work on the subtlety of expression in annotating and indexing remains for the future. When the coder of a document 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. He could complete or recite the basic sentence of our formula, supplying the interrogatives with answers. But he can only lend part of his appreciation of the document to the system for the final combining, permuting and printing-out. The syntax, and therefore much of the meaning, is lost in the successive transformations of the material thought.¹¹ J.C. Gardin and others have worked on the problem, 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 print-outs 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 cooperating 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 1200-page book as compared with a 400-page book. Or, let us guess, between \$40 and \$35.

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. Professors may order for their 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 fulfillment 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 set-up 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 listing 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.¹²

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 bibliography. 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.

FOOTNOTES

¹Alfred de Grazia, "The Universal Reference System," VIII The American Behavioral Scientist (April 1965), 3-4.

²Cf. H. Lasswell and A. Kaplan, Power and Society, (New Haven: Yale University Press, 1950).

³International Affairs, Volume I of the Political Science, Government & Public Policy Series of the Universal Reference System (80 East 11 Street, New York, N.Y. 10003: Universal Reference System, 1965) 1205 pp.

⁴At the conclusion of a recent book on Some Fundamentals of Information Retrieval (London: House and Maxwell, 1965), the author, John R. Sharp declares: "if we could reach the position wherein we had a vocabulary of terms which could be synthesised to meet any particular requirement in subject specification, and if there was only one way in which those terms could be used to provide any such subject specification then such a vocabulary used with concept coordination for the purpose of providing maximum flexibility in manipulation would solve the retrieval problem completely." (Well, Mr. Sharp, to quote the Gilbert & Sullivan General, "almost completely".)

⁵The Institute for Computer Research in the Humanities at New York University publishes an ICRH Newsletter (1965ff.) that follows this rapidly developing area. The American Documentation Institute publishes a new bibliography devoted to information retrieval that is of the highest calibre. See also Ralph L. Bisco, "Social Science Data Archives," LX APSR #1 (Mar. '66), 93-109), and Wm. A. Glaser, Plans of the Council of Social Science Data Archives, (49pp. mimeo., CSSDA, 605 West 115th Street, New York, N.Y. 10025) Cf. "Information Retrieval in the Social Sciences," special issue of The American Behavioral Scientist, Vol. VII (June 1964), 1-29, 45-71).

⁶D.J. Foskett, Classification and Indexing in The Social Sciences (Washington: Butterworths, 1963).

⁷Cf. Chris Argyris, Organization and Innovation (Homewood, Illinois: R.D. Irwin, Inc., 1965) who says: "researchers tend to be incompetent in interpersonal aspects of their work, even as their technical competence increases." Also R. Lippitt, et. al., The Dynamics of Planned Change: the authors postulate (unconsciously) a resister of change who is the villain of the book.

⁸Most importantly the studies of the Project on Scientific Information Exchange in Psychology, American Psychological Assn., under the direction of William D. Garvey and Belver C. Griffith.

⁹For a note on this unresearched subject, see Derek J. de S. Price, "Networks of Scientific Papers," 149 Science (1965), 510-11.

¹⁰The educational publishing industry, one of the fastest-growing in the nation, is in the throes of reorganization and innovation. As examples of what has happened recently, the following mergers and agreements may be cited: RCA and Random House; GE and Time, Inc., ITT and Bobbs-Merrill (Howard Sams); Raytheon and D.C. Heath; IBM and Science Research Associates; Xerox and University Films; GTE and Reader's Digest. Cf. The New Republic articles by James Ridgeway, June 4, and David Dempsey, May 14. Cf. Alfred de Grazia, ed., Social Invention, a special issue of The American Behavioral Scientist, V (Dec. 1961). Also A. de Grazia, "Social Invention in the Age of Controls," IV ABS (Oct.'60), 36-38.

¹¹J.C. Gardin, Syntol (Graduate School of Library Service, Rutgers University: New Brunswick, New Jersey, 1965).

¹²Cf. H. Borko, The Conceptual Foundations of Information Systems, (Santa Monica, California: Systems Development Corporation, May 6, 1965) 37pp. and Noah S. Prywes, "Browsing in an Automated Library through Remote Access," pp. 105-130 in M.A. Sass and W. D. Wilkinson, eds., Computer Augmentation of Human Reasoning (Washington, D.C.: Spartan, 1965)