# The Who and Why of ISI®

In a modern office building, a stone's throw from Philadelphia's Independence Hall, some 250 associates of the INSTITUTE FOR SCIENTIFIC INFORMATION are busy providing services and turning out publications specifically designed to help scientists cope with their literature and information problems. In addition to providing these services, now used by more than 150,000 scientists, ISI devotes a large percentage of its budget to research and development of new techniques for handling scientific information.

ISI's pioneering research has nurtured one of the most promising new concepts for information retrieval developed during the past 50 years—the concept of *citation indexing*.

## SCIENCE CITATION INDEX

The SCIENCE CITATION INDEX,® a comprehensive, multidisciplinary Index to science, is based on this concept, and is published quarterly and annually by ISI. The SCIENCE CITATION INDEX enables librarians, administrators and scientists to retrieve both scientific and technological information in a manner heretofore not thought possible. Although the SCI® system has been in current operation for only five years, it has gained wide acceptance by academic, industrial and government researchers.

For 1968 alone, approximately 350,000 source items were indexed to generate the SCI for that calendar

Essays of an Information Scientist, Vol:1, p.33-37, 1962-73
Current Contents, #4, March 25, 1969
Reprinted from: Karger Gazette, (13) p.2, March 5, 1966

year. This involved the processing of some 4,000,000 cited references originally written by approximately one-third of a million authors.

## Basic Concept

The basic concept of citation indexing is based on the observation that when one article cites another article there must be a subject relationship between the two articles. Each bibliographic citation as, e.g., 'A. Einstein, Ann. Physik *19*, 289 (1906)', is an unambiguous descriptor or symbol which designates the subject matter discussed in some *aspect* of the citing work. For this reason, citation indexing provides the scientist a unique and efficient method for conducting, in a minimum of time, highly specific searches, since each cited reference becomes an entry point in the Index.

The SCIENCE CITATION INDEX enables one to trace the literature forward in time; that is, to go from an earlier *cited* article to a later *citing* article. This contrasts with conventional systems in which one usually locates a current article and builds up a bibliography by tracing backward in time through the footnotes cited in the article at hand. SCI indexing is based on the author's rather than the indexer's estimate of the subject content of an article. For this reason, the SCI has been said to be particularly responsive to the user's, rather than the indexer's viewpoint. For the time being, SCIENCE CITATION INDEXES are available only for the years 1961, 1964 to 1968. The 1969 Index is, of course, now in production. ISI plans to publish Indexes for 1962 and 1963 in 1970. In addition, discipline-oriented decennial indexes in chemistry, physics, mathematics, biology, etc. are also scheduled.

## Expanded Services

During the past ten years, ISI has initiated such services as CURRENT CONTENTS®/Life Sciences, CURRENT CONTENTS/Physical Sciences, CURRENT CONTENTS/Education and CURRENT CONTENTS/Behavioral, Social and Management Sciences, which are estimated to be read by over 100,000 scientists and scholars each week. ISI introduced the INDEX CHEMICUS® in 1960, a unique abstracting-indexing service that pinpoints and reports graphically on approximately 160,000 newly synthesized chemicals each year. The INDEX CHEMICUS is estimated to be read regularly by 10,000 synthetic chemists, pharmacologists and others throughout the world—at least 50% outside the United States.

In 1968, the first commercially available computerized chemical registry system was announced by ISI. The INDEX CHEMICUS REGISTRY SYSTEM® provides both retrospective and current search capability of chemical sub-structures. ICRS™ is provided in both printed and magnetic tape form. RADICAL™ is a software system developed by ISI for use with ICRS.

In 1965, ISI introduced the AUTOMATIC SUBJECT CITATION ALERT. ASCA® is, in part, based on the citation indexing principle and is the first large scale computer-based service commercially available for selective dissemination of information to individual scientists and/or groups that covers all major disciplines.

## ISI's Philosophy

One of the major precepts of ISI's philosophy is that the various fields of science and technology are, in fact, interdisciplinary and should not be categorized into arbitrary *a priori* classifications. Many of the world's literature abstracting organizations continue to operate on an *a priori* basis, but ISI believes that it is not proper to try to rigidly define the perimeters of fast-changing areas of research. Highly interdisciplinary fields such as molecular biology,

oceanography, behavioral science, etc., require that information be extracted simultaneously from the literature of all disciplines.

ISI is attempting to *broaden* the *base* of coverage of all its services while at the same time making it easier for the individual scientist to *narrow* down and extract that *facet* of multidisciplinary information of specific interest and value to him.

It has been estimated from various sources that some 50,000 scientific and technical journals are being published in more than 65 languages. Estimates from ISI indicate that less than 2,000 journals account for over 90% of the significant scientific reports—approximately 250,000 articles per year—of which a considerable portion is duplication in various forms. This means that, contrary to popular belief, a relatively small number of journals account for a high percentage of the significant articles published. ISI's present journal coverage already exceeds 2,500 journals.

## Processing

While the processing of several hundred thousand items each year is no trivial problem, ISI's studies have led to both manual and computer systems that enable the scientist to overcome the sense of helplessness that is conveyed by unqualified allusions to 50,000 journals—many of which ceased publication long ago, or publish a trivial number of significant papers. While there has been a steady growth of the literature, ISI believes the so-called "information explosion" can indeed be controlled. Contrary to the general cliché that "there is too much scientific information", ISI would assert that "there is often little or no information available about a specific problem". This "needle in the haystack problem" is faced almost daily by every scientist. In many such cases, conventional literature searching techniques may not be practical. However, such techniques as the SCIENCE CITATION INDEX and ASCA enable the searcher to find out quickly if there is—or is not—information available on a specific point. By reducing the time required to do comprehensive but specific searches, ISI hopes to increase the use of the literature as an effective research instrument.

Information science as practiced by conventional libraries and information centers has not usually been thought to be an important or exciting part of research and development by scientists. However, with the increased opportunities for productive exploitation of the scientific literature, this situation is now changing. Computer technology has also had great impact on the field, making large scale information handling systems practical. More than half of ISI's total working force consists of information scientists, librarians and data processing technicians. An IBM 360 system is used to process about 150,000 new punch cards per week. And the cards are then converted to magnetic tape. Using special programs developed at ISI, these tapes are scanned automatically for scientists all over the world. ISI's magnetic tapes and software are provided in a variety of forms on leasing arrangements with foreign governments and institutions in Canada, the United Kingdom, Sweden, France, etc. ISI also maintains offices in Europe and Japan.

## ISI's Goal

ISI's all-consuming goal is to organize the world's total output of significant scientific and technical literature into an integrated file and provide convenient current and retrospective access to the file by individual scientists. By the end of 1969, ISI will have processed approximately 15 million reference records. The goal of total retrospective coverage for the literature of the twentieth century should be completed in gradual stages over the next decade. This

enormous file will be of inestimable value both for current laboratory research and also for new types of historical studies in which ISI has pioneered.

## Research Program

While the file building is going on, ISI's research program is experimenting with new techniques on the commercial services already being provided by ISI. The fruits of this research already have started to pay off in the new ISI SEARCH SERVICE in which it is possible to obtain a retrospective search of the literature on a specific topic using ISI's data bank. Searches are performed daily at low cost—often by telephone or cable request.

While ISI's initial program was geared to the so-called natural and physical sciences, the social and behavioral sciences have not been forgotten. The heightened interest in these fields is reflected in CURRENT CONTENTS/Behavioral, Social and Management Sciences as well as in CURRENT CONTENTS/Education. These new services will accelerate ISI's already heavy coverage of the behavioral sciences in its SCIENCE CITATION INDEXES and PERMUTERM® SUBJECT INDEXES. They are also the harbinger of the SOCIAL AND BEHAVIORAL SCIENCES CITATION INDEXES that ISI plans to publish both currently and retrospectively in the days ahead. This means that scholars in all fields can now benefit from ASCA service on a current basis and use the revolutionary principle of citation indexing to add a new dmension to scholarship.

## PERMUTERM SUBJECT INDEX

In 1966, ISI established another first by the introduction of its PERMUTERM SUBJECT INDEX. Based on the natural language of article titles and augmented by translation of foreign titles into English the ™PSI's computer is then used to create all possible permuted pairs of terms. These co-occurrences of terms facilitate the search process while maintaining a constant, self-correcting contact with actual literature usage. The PSI itself is an invaluable source of terminological data useful in constructing search profiles and augments the use of the CITATION INDEX when key references are not known or are forgotten. Just as the SOURCE INDEXES of the SCI can be used to identify citations by author, the PSI can be used to identify papers from key-words or partially recalled titles.

## Magnetic Tapes

Even at the present time, ISI is supplying copies of its magnetic tapes to various organizations for their internal use. These tapes are used in many ways including citation and word-searching approaches to subject matter retrieval. A new approach which simplifies the complexities of existing word-profile schemes has been fully tested and debugged. This additional approach augments the ASCA citation profiles and enables scientists to construct word profiles to search literature for concepts that do not lend themselves as readily to the citation approach. The combined approaches through words, authors, organizations, citations, journals and stems will provide a depth of indexing to satisfy the most exacting requirements at low cost.

Studies on the sociological implications of science and scentific literature are an important part of ISI's R & D program. A recent report for the US Government on *The Use of Citation Data in Writing the History of Science* is looked upon by many in the field as having unique application to pinpointing areas for future scientific research, identifying the research fronts of science, and evaluating published work by providing convenient access to criticisms by peers.

## OATS Service

Supporting all of ISI's retrieval and dissemination activities is an original system of storing scientific journals. ISI learned from the outset that it must back up any bibliographic listing in CURRENT CONTENTS, INDEX CHEMICUS, etc., with convenient access to papers listed. It is generally expected that the scientist will use local library facilities for this purpose. In addition, reprints are frequently requested directly from authors by using the computer-produced address directories which ISI publishes in CURRENT CONTENTS and in the INTERNATIONAL DIRECTORY OF RESEARCH & DEVELOPMENT SCIENTISTS.™ Nevertheless, there is frequently the need for the scientist to obtain quickly a paper that is not received at his institution. For this reason, ISI operates the ORIGINAL ARTICLE TEAR SHEET service. In the OATS® service, tear sheets are literally torn from extra copies of journals received by ISI. Although many requests are received at ISI, only a small number of requests are for any given article. This enables the Institute to satisfy most needs from a maximum of five or six copies of each issue. ISI has also signed copyright agreements with leading publishers throughout the world to whom royalties are paid for tear sheets and photocopies.

ISI takes great pride in the good relations it has maintained with hundreds of publishers throughout the world who cooperate to minimize the many inherent time lags in publication and recognize that the growth of interdisciplinary science requires that individuals have access to many journals they do not ordinarily scan directly. The growth in popularity of ISI's services has been paralleled by a comparable increase in journal circulation throughout the world.

## Systematic Approach

The key to effective retrieval of information is to convert the process of information discovery to that of information recovery. By effectively disseminating information in the first place, as through CURRENT CONTENTS, INDEX CHEMICUS, and ASCA, the memory of the individual scientist can be better exploited to recover that information when it is needed again. On the other hand, since the individual scientist can only be exposed to a fraction of the world's information in his lifetime, he also needs what has been described as "systematic serendipity"—an organized process of information discovery of that which he did not know existed. Finally, he must be able to quickly and efficiently determine if it is valid to conclude that *no* information is available and he can proceed in the laboratory with a reasonable degree of assurance that unwitting duplication of efforts will be avoided.

The SCI and ASCA serve these two latter functions. The above objectives of information science indicate the important creative and service roles the entire information processing profession can play in future research and development activity throughout the world.

### *References*


1. GARFIELD, E.: Science Citation Index. A new dimension in indexing. *Science 144*: 649-654 (1964).
2. GARFIELD, E.; SHER, I. H.: ASCA (Automatic Subject Citation Alert)—A new personalized current awareness service for scientists. American Behavioral Scientist *10*(5):29-32 (1967).
3. GARFIELD, E.: Primordial concepts, citation indexing and historio-bibliography. Journal of Library History *2*(3):235-249 (1967).
4. MALIN, M. V.: The Science Citation Index —A new concept in indexing. Library Trends *16*(3):374-387 (1968).
5. GARFIELD, E.: The information implosion. Chemistry *41*(7): 24-31 (1968).
6. GARFIELD, E.: Citation indexing: A natural science literature retrieval system for the social sciences. American Behavioral Scientist *7*:58-61 (1964).
7. GARFIELD, E.; SHER, I. H. and TORPIE, R. J.: The use of citation data in writing the history of science (Philadelphia: Institute for Scientific Information 1964), p. 75.