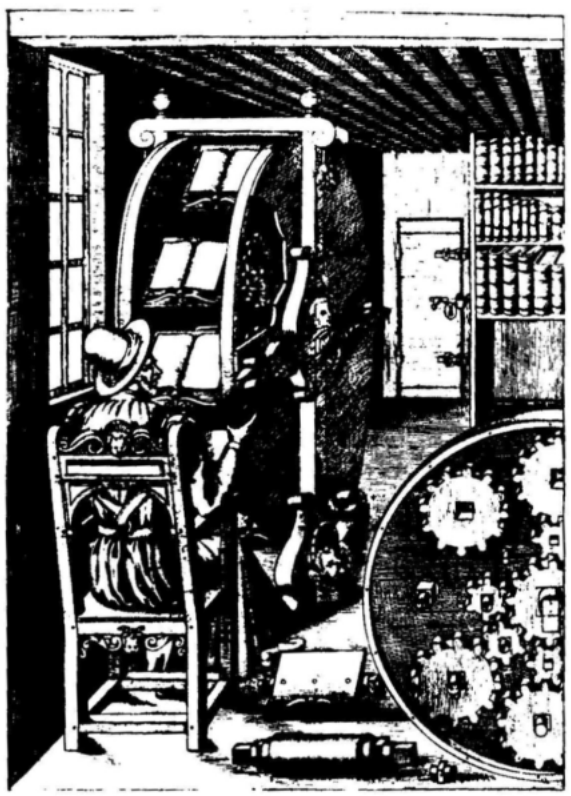
Garfield, E. "The information implosion ," Chemistry, 41 (7) p.24-31, 1968.

THE INFORMATION IMPLOSION

Eugene Garfield
Institute for Scientific Information
Philadelphia, Pa.

Before Newton's time, scientific information was disseminated largely by personal correspondence; by Franklin's time, the journal had come into common use; and by 1830, the number of journals had grown to the point where abstracting service was needed. Today, with the volume of scientific information doubling every 10 to 15 years, high speed computers are needed, and a whole new field of information science has sprung up, with several active organizations.



An earlier solution to the information problem

One such organization is ISI (Institute for Scientific Information) in Philadelphia, Pa., where experts think in terms of an information implosion. The purpose of the organization is to enable scientists to use the vast amount of scientific information now available and work is based on three concepts:

—Despite the huge number of scientific journals published today, most of the important information can be found in a relatively small number of core journals.

—All science has become so interrelated that attempts to restrict an information service to one branch of science would result in the loss of much valuable information.

—Searching methods other than by reference to subject and author might be equally valuable.

Core Journals Concept

Although the total number of scientific journals published throughout the world is claimed to be about 35,000 to 100,000, studies made by ISI indicate that fewer than 1000 journals account for over 90% of the significant scientific reports. This is certainly still too many for an individual alone to scan. But it is not too many to be covered by an organization such as ISI, which at present processes more than 2000 journals or, in other words, well over 90% of the significant scientific information published throughout the world. Further, ISI's yearly increase in coverage has so far exceeded the general growth rate of science. Thus, the gap is steadily narrowing between existing coverage and total or ideal coverage, which is probably unattainable.

The Network of Science

Most information services attempt to cover one field or one discipline. ISI, however, feels that

science and technology have become one vast cross-disciplinary network. Fields such as molecular biology, oceanography, environmental science, and aerospace science cannot be placed into neat little cells. To be reasonably sure that he has found most of the important information in any such field, an individual must screen a substantial percentage of all the significant journals of science. For example, a chemist involved in the problems of selecting materials suitable for use in surgical implants or artificial internal organs might find information of interest in publications ranging from *Materials Research and Standards*, an engineering journal, to the *Journal of Cardiac Surgery*, a medical journal, to a chemical journal such as the *Journal of Polymer Science*.

Standing on the Shoulders of Giants

Probably the most interesting of ISI's findings is the utility of citation indexing in gaining access to the literature of science. August Kekulé, the famous German chemist, stated at the twenty-fifth anniversary of his benzene structure that we are able to make progress in science because we stand on the shoulders of giants. The statement can be extended by saying that all scientists stand on each others' shoulders. Citation indexing is based on this idea. If we know that the basic work on stereospecific polymers was done by Giulio Natta, then we can find what has been done since by looking to see who is standing on Natta's shoulders.

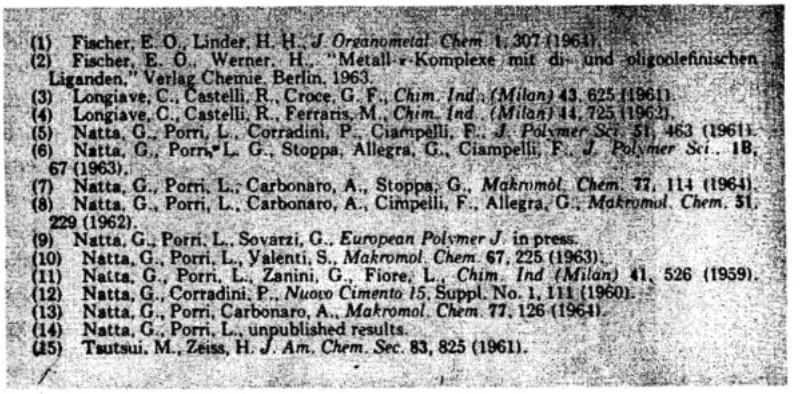
A mechanism for doing just that is built into the formal structure of the scientific paper. The author of a paper is required to acknowledge prior work upon which his paper is based. These acknowledgments usually take the form of citations or references giving the name of the author and the volume, page, and year of the journal in which the paper appeared (Figure 1). The purpose of the citation is twofold. It gives credit for earlier work but, most important, it enables the reader to obtain additional details.

A citation index does something more. It enables one to find out who is standing on whose shoulders. If we want to know what has happened in polymer chemistry based on Natta's work, we can do so by obtaining a list of all papers citing Natta's basic paper. The men who wrote those papers are the men who are working in the area in which we are interested, and the papers in which they have cited Natta's work are papers we should see. This is the essence of citation indexing.

Services to Chemists

Services of ISI that are of particular interest to chemists include all three editions of *Current Contents, Index Chemicus,* the *Science Citation Index*, and *Automatic Subject Citation Alert (ASCA III),* and the *ISI Search Service.*

Current Contents. *Current Contents (CC)* is a weekly alerting service in which the contents



(1) Fischer, E. O., Linder, H. H., *J. Organometal. Chem.* 1, 307 (1964).
(2) Fischer, E. O., Werner, H., "Metall-π-Komplexe mit di- und oligoolefinischen Liganden," Verlag Chemie, Berlin, 1963.
(3) Longiave, C., Castelli, R., Croce, G. F., *Chim. Ind. (Milan)* 43, 625 (1961).
(4) Longiave, C., Castelli, R., Ferraris, M., *Chim. Ind. (Milan)* 44, 725 (1962).
(5) Natta, G., Porri, L., Corradini, P., Ciampelli, F., *J. Polymer Sci.* 51, 463 (1961).
(6) Natta, G., Porri, L. G., Stoppa, Allegra, G., Ciampelli, F., *J. Polymer Sci.*, 1B, 67 (1963).
(7) Natta, G., Porri, L., Carbonaro, A., Stoppa, G., *Makromol. Chem.* 77, 114 (1964).
(8) Natta, G., Porri, L., Carbonaro, A., Cimpelli, F., Allegra, G., *Makromol. Chem.* 51, 229 (1962).
(9) Natta, G., Porri, L., Sovarzi, G., *European Polymer J.* in press.
(10) Natta, G., Porri, L., Valenti, S., *Makromol. Chem.* 67, 225 (1963).
(11) Natta, G., Porri, L., Zanini, G., Fiore, L., *Chim. Ind (Milan)* 41, 526 (1959).
(12) Natta, G., Corradini, P., *Nuovo Cimento 15*, Suppl. No. 1, 111 (1960).
(13) Natta, G., Porri, Carbonaro, A., *Makromol. Chem.* 77, 126 (1964).
(14) Natta, G., Porri, L., unpublished results.
(15) Tsutsui, M., Zeiss, H. *J. Am. Chem. Soc.* 83, 825 (1961).

Figure 1. The author of a scientific paper is required to acknowledge prior work on which his paper is based

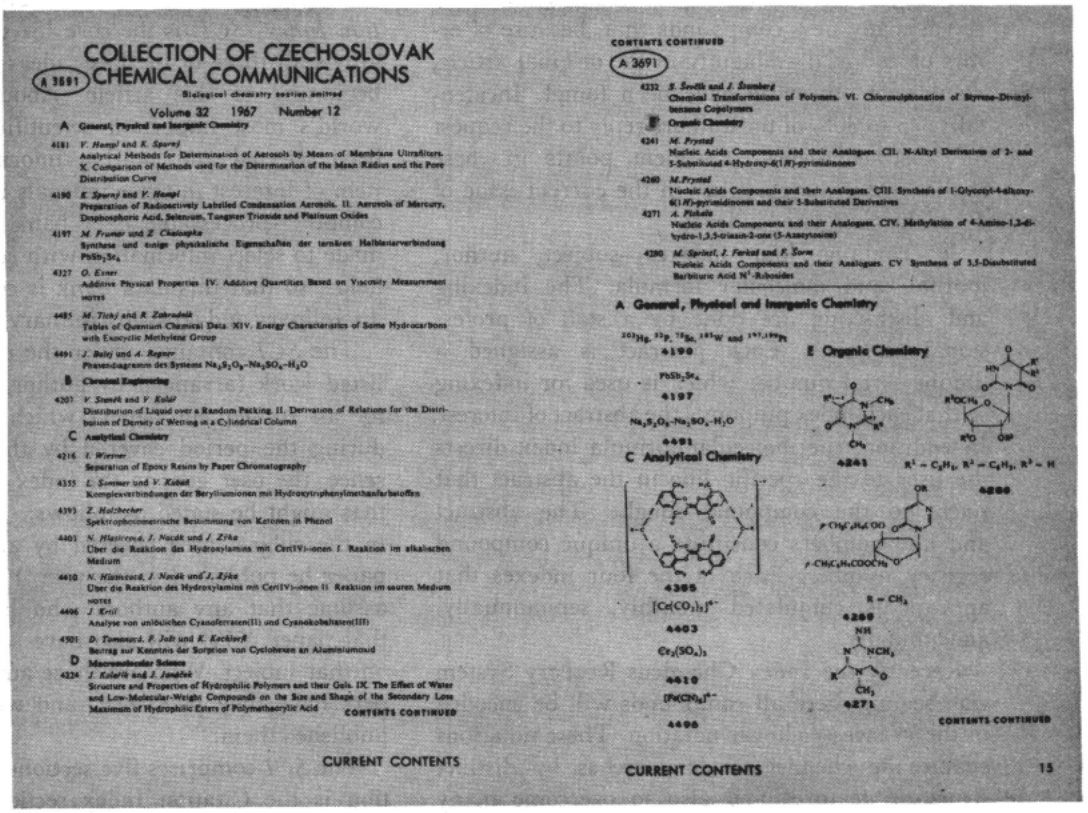
pages of a large group of journals are reproduced. This publication enables the scientist to have, in effect, hundreds of journals pass over his desk each week for scanning. Three editions of *CC* are published: *Life Sciences, Physical Sciences,* and *Chemical Sciences.* Taken together, these three publications cover the world's significant scientific journals. All three editions have heavy coverage of chemistry. *Life Sciences* stresses biochemical research; *Physical Sciences* stresses applied, inorganic, and physical chemistry. Most of the same basic chemical journals appear in all *CC* editions.

CC Chemical Sciences (CCCS) was designed specifically for the industrial chemist. It reprints, on a weekly basis, the tables of contents of a carefully selected group of key foreign and domestic chemical journals. Also, it includes structural diagrams, when available, taken from the original publication (Figure 2). All three of the *CC* publications include an author index and address directory. But *Chemical Sciences* has, in addition, a detailed subject index prepared by a unique combination of man and machine. Chemically trained indexers select key terms which are then permuted by an IBM computer to generate all possible permuted pairs. This is called the Permuterm Index. Here all terms assigned to a given article are formed into all possible pairs of terms—that is, all possible permutations of N terms taken two at a time. This produces $N(N-1)$ permuted pairs per article.

Index Chemicus. *Index Chemicus (IC)* is a weekly service that reports on the new chemical compounds appearing in the world's journals. It is, in fact, a registry of new chemical compounds. The service is especially oriented to the needs of the chemist involved in organic and inorganic synthesis. In 1967, *IC* reported on 147,000 new compounds; this year the total is expected to exceed 165,000.

In most cases, *IC* reports the new compound within 30 to 60 days of the receipt of the pertinent journal. This not only includes the time required for author approval but also printing and distribution. The publication arrives in Europe and Japan at about the same time it is received in California.

In addition to complete bibliographic information, the abstract shown in Figure 3 includes the author's summary of the content of the paper, structural formulas and flow diagrams, a complete list of the molecular formulas of all compounds which the authors claim to be new or which are presumed to be new by the ISI staff. If the original article indicates a use for the compound, that use is noted under "Use Profile." The *IC* abstract also includes an instrumental


COLLECTION OF CZECHOSLOVAK CHEMICAL COMMUNICATIONS

A 3691

Volume 32 1967 Number 12

CONTENTS CONTINUED

CURRENT CONTENTS

Figure 2. Pages from Current Contents, Chemical Sciences. Tables of contents for a selected group of key foreign and domestic chemical journals are published weekly

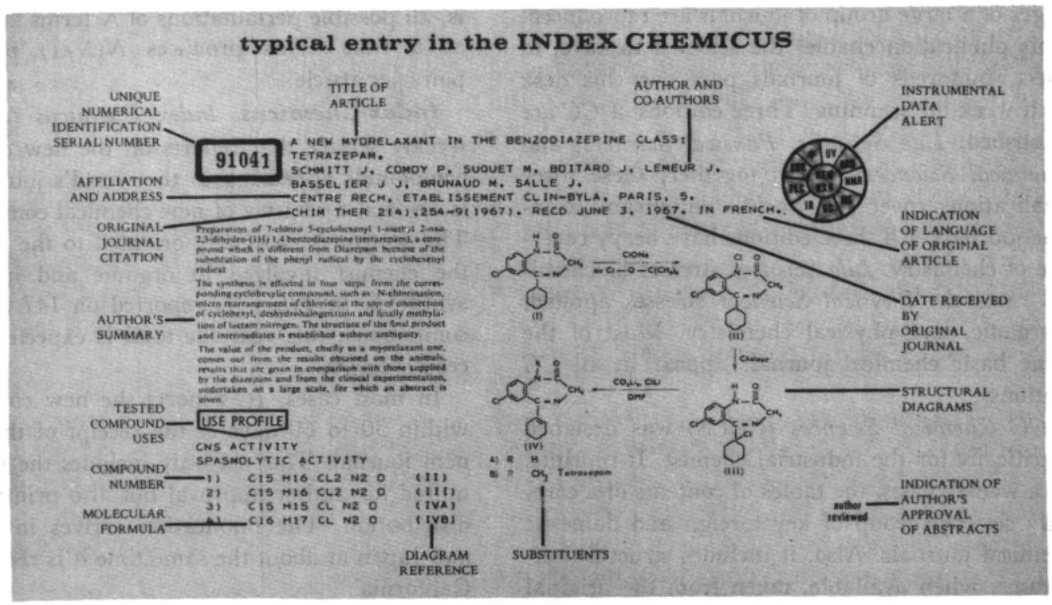


Figure 3. Index Chemicus is a weekly publication that reports new chemical compounds appearing in the world's journals

data disk which indicates new reactions and information such as ultraviolet, infrared, and nuclear magnetic resonance spectra.

Another unique feature of *IC* is author review of the abstracts. Each abstract is sent to the author for review before publication, and this is indicated at the lower right of the abstract. An interesting effect of this author review has been to reveal many new compounds that, because of errors or editorial ambiguity in the original article, would not otherwise have been found. Incidentally, 75 to 80% of the authors reply to the request for abstract review. Pertinent points in these replies are included either in the current issue of *IC* or the cumulated indexes.

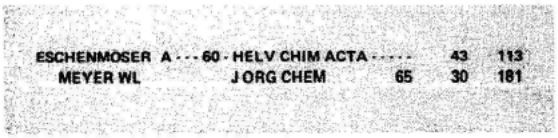
IC contains four indexes—subject, author, journal, and molecular formula. The indexing and abstracting are done by a staff of professional chemists. Each abstract is assigned a unique serial number which is used for indexing so that each index pinpoints the abstract of interest. In addition, the molecular formula index directs the user to the specific line in the abstract that refers to the compound sought. The abstract and line numbers constitute a unique compound registry number. Each of the four indexes that appears is cumulated monthly, semiannually, and annually.

Recently the Index Chemicus Registry System was begun where all compounds will be encoded in the Wiswesser linear notation. These notations enable the chemist to do searches by distinct structural features and also to overcome many difficulties associated with traditional chemical nomenclature. The notations are supplied on magnetic tapes which can also be used to provide special services such as selective dissemination. Thus, a chemist working on a given class of chemicals with special structural characteristics could be regularly informed of any new compounds in the literature containing these structural features.

Science Citation Index. The *Science Citation Index (SCI)* is the core ISI service that integrates all three of the basic ideas discussed at the beginning of this article. About 2000 of the world's most important scientific journals have been selected and citation information on each item of interest in these journals is entered into a computer-processable data bank. No attempt is made to select information with regard to specific fields, so that this data bank is completely interdisciplinary and cross-disciplinary.

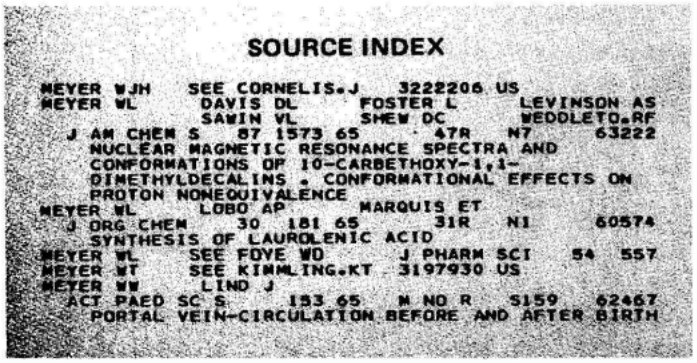
The *SCI* contains, under the citation for each listed work (arranged by author), the references for all subsequent papers which cited the work during the period covered by the index. In essence, the user goes to the index with a question that might be stated as follows: "I am interested in the subject area covered by author *X* in the paper he published in the year *Y* in journal *Z*. I assume that any authors who have since cited that paper *XYZ* will also have something to say on that subject. Who are those authors, what papers have they published, and where have they published them?"

The *SCI* comprises five sections. The main section is the Citation Index section; in addition, there are the Source Index, Corporate Index, Patent Index, and Permuterm Index sections. In



ESCHENMOSER A ··· 60 · HELV CHIM ACTA ····· 43 113
MEYER WL J ORG CHEM 65 30 181

Figure 4. Typical entries in the Science Citation Index. Articles cited during a specified period are listed alphabetically according to first author's name. Then each work citing that author, indicated by a dashed line, is listed chronologically



SOURCE INDEX

MEYER WJH SEE CORNELIS.J 3222206 US
MEYER WL DAVIS DL FOSTER L LEVINSON AS
SAWIN VL SMEW DC WEDDLETO.RF
J AM CHEM S 87 1573 65 47R N7 63222
NUCLEAR MAGNETIC RESONANCE SPECTRA AND
CONFORMATIONS OF 10-CARBETHOXY-1,1-
DIMETHYLDECALINS . CONFORMATIONAL EFFECTS ON
PROTON NONEQUIVALENCE
MEYER WL LOBO AP MARQUIS ET
J ORG CHEM 30 181 65 31R N1 60574
SYNTHESIS OF LAUROLENIC ACID
MEYER WL SEE FOYE WO J PHARM SCI 54 557
MEYER WT SEE KINMLING.KT 3197930 US
MEYER WW LIND J
ACT PAED SC S 153 65 M NO R 5159 62467
PORTAL VEIN-CIRCULATION BEFORE AND AFTER BIRTH

Figure 5. The Source Index compiles, under the name of the first author, title, coauthors, and bibliographic data for each cited item

the Citation Index, any article cited during the period indexed is listed alphabetically under the first author's name. Each cited work by that author is arranged chronologically and indicated by a separate dashed line. Beneath each cited item is the list of current citing articles.

In the example shown in Figure 4, the 1960 article by Eschenmoser in *Helvetica Chimica Acta,* volume **43,** page 113, was cited by W. L. Meyer in the *Journal of Organic Chemistry,* 1965, volume **30,** page 181.

The Source Index (Figure 5) lists, under the name of the first author, the full title, coauthors, and bibliographic data for each citing item. To put it another way, all the current items published in the period indexed are listed alphabetically by author. Each item is coded, as in the Citation Index, to indicate its identity such as a correction, a book review, a meeting, or a letter. The Source Index is essentially a giant calendar-year author index; that is, it includes all the journal items processed by the closing date for the current year. The 1967 Source Index covers over 95% of the journal issues of that year. Thus, all 1967 issues of *Chemistry and Industry* appear in the 1967 index. The Source Index can, of course, be used independently of the rest of the *SCI.*

Although the Citation Index is essentially a subject index used to locate where a given paper has been recently cited, the Source Index is the comprehensive author index of science, used to locate what a given author has published during a particular year. The Corporate Index provides a similar function for organizations such as a university chemistry department or the research division of a chemical firm.

The *Permuterm Subject Index (PSI)* provides the subject word approach as described previously for *CCCS,* but the *PSI* covers all 300,000 source items indexed.

Automatic Subject Citation Alert. *ASCA III* is a system based upon the *Science Citation Index* data bank for alerting the scientist to the publication of new articles within his field of interest. Like *SCI, ASCA III* also covers close to 2000 key journals; each week, approximately 100,000 references and 7500 source items are added to the data bank from the current journal literature. *ASCA III* notifies the subscribing scientist of items that will be of specific interest in his work. This notification system is based upon a profile of interests submitted by the subscriber. An *ASCA III* profile will usually consist, in part, of a list of specific papers and a list of authors. The client will be notified any time that a newly published article cites any paper or author in his profile.

In addition to the citation-based questions, the

The title, "Role of Impurities in Thermal Ionization" has 3 significant words (N=3). When indexed, this produces 3(3-1) or 6 permuted pairs as follows:
Impurity-Ionization
Impurity-Thermal
Ionization-Impurity
Ionization-Thermal
Thermal-Impurity
Thermal-Ionization

subscriber may also include requests for articles which contain any specific word, word stem, or phrase in their titles; are written by a given author; describe work done at a given organization; or are published in a given journal. In framing his questions in the above terms, the scientist may also specify combinations of the four types of items. For example, one might ask to be notified of any article by James Jones that contains the term catecholamines in the title and was written while Jones was at Harvard Medical School, provided that the article was found in a journal other than the *Journal of the American Chemical Society* (since the client regularly scans this journal himself).

Figure 6 shows a typical weekly *ASCA III* report. It includes complete bibliographic data for each journal item listed, and an indication as to which profile term caused the item to be listed.

ISI Search Service. The ISI retrospective search service is primarily based on the *SCI* data bank, though it can be augmented by traditional methods at the user's request. The using scientist may request that he be informed of any item that cited a particular article, book, or author during a specified period. He may also ask for a list of items published, during the specified period, by a given author or attributed to a specific organization. Naturally, he may also request a search of the *Permuterm Subject Index*. For the chemist there is also a similar search service based on the *Index Chemicus* files of approximately 900,000 compounds.

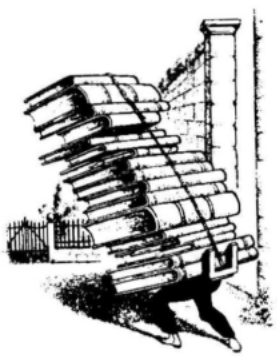
The Chemist's Information Problems

The chemist's specific information needs depend on the field he is in and on the particular research project in which he is involved. Nevertheless, regardless of the type of research—whether it be synthesis, analysis, structure identification, reaction mechanisms and kinetics, the isolation of natural products, or biological activities—a number of generalizations can be made. Any chemist requires information on two levels: First, he must remain reasonably conversant with the over-all field of chemistry; second, he requires detailed information related to his current research projects. One might characterize these needs as being, respectively, extensive and intensive.

Extensive Information Needs. By scanning *CC* publications, the chemist can quickly obtain an overview that will tell him where the major research fronts are and enable him to obtain details in areas of significant interest to him. For the industrial chemist, *CCCS* provides a more sharply focused service that is particularly designed for his use. In a similar manner, *Index*

```
a s c a
AUTOMATIC SUBJECT CITATION ALERT

a service of the INSTITUTE FOR SCIENTIFIC INFORMATION

   DR. BARON H. TRISSEN                          D6527 ACCOUNT NUMBER
   R+D CONSULTANTS                                  56 $ UNITS USED
   BOSTON, MASSACHUSETTS                            44 $ UNITS UNUSED

                                                 51634      D6527

   REPORT FOR   7 OCT 66

180,436 INDEXING TERMS FROM CURRENT SCIENTIFIC LITERATURE
        WERE PROCESSED FOR ASCA THIS WEEK

 THE ITEM BY CANTONE A       INT Z ANGEW PHYSIOL       18  107 60
 THE ITEM BY HALONEN PI      NATURE                    93  942 62
  CITED BY  WEGMANN HM   BRUNER H    KLEIN KE     VOIGT ED
            FED PROC      25 1405 66  M  37R   N4P1   81860
 HORMON/      ENZYMATIC AND HORMONAL RESPONSES TO EXERCISE
              LOWERED PRESSURE AND ACCELERATION IN HUMAN
              PLASMA AND THEIR CORRELATION TO INDIVIDUAL
              TOLERANCES

  ORGANIZATION CZECH AC SCIENC I SOL STAT PHYS PRAGU
  PUBLISHED DORDA G       VRBA J
            PHYS ST SOL   17 K127 66       2R`  NI     81899
              LONG-TIME EFFECT OF EXTERNAL FIELDS ON SURFACE
              SILICON OXIDE CONDUCTIVITY

 (SOURCE AUTH) ROSS BB
   PUBLISHED DANIEL SS   DAWES GS      JAMES LS    ROSS BB
             BR MED J      2  562 66        8R  N5513  81722
               ANALEPTICS AND RESUSCITATION OF ASPHYXIATED
               MONKEYS

 THE ITEM BY COLBY KM          ARCH GEN PSYCHIAT      10  220 64
 THE BOOK BY FREUND JE         MATHEMATICAL STATIST         62
  CITED BY  NAYLOR TH    GIANTURC.DT
              ARCH G PSYC   15  293 66      26R   N3     82213
                COMPUTER SIMULATION IN PSYCHIATRY

 REF AUTHOR  JANSEN LL         WEEDS                   9  381 61
 REF AUTHOR  JANSEN LL         J AGRIC FD CHEM        12  223 64
 ORGANIZATION U CAL DEP BOT DAVIS
 PUBLISHED SMITH LW      FOY CL        BAYER DE
             WEED RES        6  233 66       14R   N3     82582
               STRUCTURE-ACTIVITY RELATIONSHIPS OF ALKYLPHENOL
               ETHYLENE OXIDE ETHER NON-IONIC SURFACTANTS AND
               3 WATER-SOLUBLE HERBICIDES

           SATTINGE.DH
             J MATH+PHYS   45  188 66       3R   N2     81794
  ANISTROPIC   EIGENVALUES OF AN INTEGRAL EQUATION IN
               ANISOTROPIC NEUTRON TRANSPORT THEORY


           GAVRILOV DV
             OPT SPECT R   20  486 66       25R   N5     82569
  LENS       CALCULATION OF AN UNCEMENTED 2-LENS SYSTEM WITH
             1 REFLECTING SURFACE
```

Figure 6. The Automatic Subject Citation Alert informs a scientist of new articles published in his field of interest



Chemicus enables the organic and synthetic chemist, the pharmacologist, and the medicinal chemist to remain abreast of the new chemical compounds that are reported in all journals of importance. And it reports on these within 30 to 60 days of their first mention in the literature.

Intensive Information Needs. Essentially, a chemist about to undertake any new research project must go through the following process: First, he must decide on a line of research that he considers to be promising. This decision will probably come out of his general background—his extensive knowledge. Second, he must determine that the specific line of research has not been followed to a conclusion (either positive or negative) by an earlier chemist. If the results of the second step indicate that this is, indeed, a valid line of research, the chemist must then obtain as much of the known pertinent information as he can. Finally, having brought himself up to date, he must remain in contact with the relevant research fronts.

The particular methods used at each of these phases will vary with the field of the research. Nevertheless, a detailed knowledge of the information services available, and of the ways in which they can be used, will enable the scientist to go through each of these phases with a minimum of lost motion.

Stanley Hopeful's New Line of Research

In this short article it is not possible to convey all the detailed information on information that should be a part of the chemist's background. Even so, consider some examples of how a chemist might make use of available ISI services.

Our chemist, Stanley Hopeful, is looking for a new line of research. He recalls having read a paper by Ziebert at the University of California, which described an unusual biological activity of a glandular extract of the adult female Japanese beetle (*Popillia japonica*). At the time the article appeared, this activity was not of great interest, but with the current ferment in the field of biological pesticides, Hopeful suspects that Ziebert's paper might be worth following up.

Hopeful must now determine whether anyone else has followed up on Ziebert's paper. If he doesn't remember when and where it was published, he can first identify the exact title and date of publication of Ziebert's paper by looking at the Source Index of the *SCI*. However, with just the name and year he can consult the Citation Index volumes of *SCI* to determine what authors subsequently cited Ziebert's paper and in what context. Stanley could also have made use of the *ISI Search Service* to do this job if he did not have the time to do it himself.

After noting all of the citations to Ziebert's paper up to the present, Stanley has found that a number of workers have followed up on Ziebert's extract. Several fractions have been analyzed, but no one has been working in the direction of a biological pesticide. At this point, he decides that the idea is worth investigating further. Now he goes to the *Cumulated Index Chemicus* and investigates the compounds that have been identified in Ziebert's extract. He finds that a considerable amount of work has been done and a few biological activities of the compounds have been identified but, again, no one is following the biological pesticide research line. After checking these new references in the *SCI* to make sure that no other work has been done, he decides to launch his project.

Now Stanley must see to it that he is kept aware of any new papers that might assist him or that might impinge on his research. One way he does this is by scanning each weekly issue of *Index Chemicus* as he receives it. He also checks the formula and subject indexes each month. In addition, he makes use of *ASCA III* to inform him of any newly published papers of interest.

He does this by using citations for the original Ziebert paper (and two later papers that branched off in an appropriate direction) as part of his profile. In addition, he submits a few word questions that will notify him of any other possibly related papers published on one or more aspects of his biological pesticide research.

Whenever Stanley finds a paper that seems particularly relevant by using any of these services, he can either write the author for a reprint or go to his local library to examine the journal. Or, if he prefers, he can use ISI's mail-order library service and obtain an *Original Article Tear Sheet.* ISI stores multiple copies of each journal and literally tears out the requested article. This document procurement service thus completes the information cycle.

Suggested Reading

(1) Garfield, E., "ISI's Service in the Design of Small-User Systems," *J. Chem. Doc.* **6,** 164–8 (1966).

(2) Garfield, E., "Science Citation Index—a New Dimension in Indexing," *Science* **144,** 649–54 (1964).

(3) Garfield, E., Sher, I., "ISI's Experience with ASCA—Selective Dissemination System," *J. Chem. Doc.* **7,** 147–53 (1967).

(4) Price, D. J. de S., "Networks of Scientific Papers," *Science* **149,** 510–15 (1965).

(5) Weinberg, A. M., *et al.*, "President's Science Advisory Committee on Science, Government, and Information (Responsibilities of the Technical Community and the Government in the Transfer of Information)," U.S. Government Printing Office, Washington, D.C.

EUGENE GARFIELD, president and director of the Institute for Scientific Information, Philadelphia, Pa., received his B.S. in chemistry and M.S. in library science from Columbia University and his Ph.D. in structural linguistics from the University of Pennsylvania. Dr. Garfield, who holds several U.S. and foreign patents on devices for selective copying and reproduction of information, has authored numerous papers on information science.

A reprint from

Chemistry

Vol. 41, July–Aug. 1968, Pages 24–31