Essays of an Information Scientist, Vol:7, p.194-201, 1984
Current Contents, #26, p.3-10, June 25, 1984

Current Comments®

Index Chemicus Goes Online with Graphic Access to Three Million New Organic Compounds

Number 26 June 25, 1984

Those *Current Contents*® (*CC*®) readers who are chemists know that we cover all the significant journals of chemistry in either of our two most widely used editions—*CC/Life Sciences* or *CC/Physical, Chemical & Earth Sciences*. For most readers of *CC*, scanning article titles is adequate for their current awareness needs. But for many scientists in organic chemistry, pharmacology, and other disciplines, there is a crucial need for browsing and searching at another level—the chemical compound. Since 1960, therefore, we have produced *Index Chemicus*® (*IC*®). At first, *IC* was a monthly, and then a biweekly, publication. But in 1970, it became a weekly listing of and index to newly reported compounds. Since each *IC* graphic record also includes the author's abstract when available, we changed its name to *Current Abstracts of Chemistry and Index Chemicus*® (*CAC&IC*®). In the past 24 years, *IC* has covered over 350,000 articles and listed over 3,000,000 new organic compounds.

It was inevitable that this data base would be accessible electronically. It has been available on magnetic tapes for many years, but with little fanfare we mounted *Index Chemicus Online*™ earlier this year and it became available this month. The file is stored on the host computer of the Paris-based online vendor Telesystemes. This organization has pioneered the development of the *Questel* and *DARC* software systems, which handle bibliographic and structural data, respectively.

Index Chemicus Online covers the same information reported in *CAC&IC* each week. Its main virtues are both bibliographic and structural. Two additional ISI® services, the *CAC&IC* 22-year Microform Cumulation, available since December 1983, and the *Chemistry Citation Index*™ (*CCI*™), which we'll introduce later this year, are designed to complement *Index Chemicus Online*.

By the end of 1984, *Index Chemicus Online* will include information on over 350,000 articles published 1962 to date. These papers first reported more than 3,000,000 new organic compounds. The inclusion of compounds reported between 1962 and 1964 is noteworthy since structural information for that period is not readily available online from any other source.

Index Chemicus Online covers over 90 percent of the new organic compounds reported worldwide in chemistry journals. The thoroughness of this service is due, in part, to ISI's staff of 23 professional chemists. During the course of a year, they carefully read every item, including footnotes and references, in more than 100 key journals. From a total of at least 35,000 articles, they prepare entries for the approximately 15,000 articles which report new compounds and synthetic methods. When preparing these entries, our chemists often contact

authors to clarify ambiguities or correct errors reported in the journals. This information then becomes part of the "record."

One of the features which *CAC&IC* readers find most valuable is the structural diagram included for each compound reported. Chemists find them to be a real time saver, since they are readily familiar with such diagrams. Users of *Index Chemicus Online* can also view these diagrams, *while online*, if they have a graphics terminal. These graphic displays overcome the ambiguity and complexity of chemical nomenclature. This was the original inspiration for *IC*.

Each *Index Chemicus Online* record contains the standard bibliographic data, including authors and their addresses, journal name, and article title. In addition, "flags" or "alerts" to scientific data reported in the original article appear in relevant records from 1968 to date. Of special interest is the flag for the analytical instrumentation used to isolate and identify the new compounds. These include infrared spectroscopy, nuclear magnetic resonance, and gas chromatography. Other alerts flag data on biological activities, explosive reactions, isotopically labeled compounds, and new synthetic methods. Experimental details regarding these synthetic methods are available in our printed publication *Current Chemical Reactions*®.[1] Table 1 lists the displayable fields in *Index Chemicus Online*.

Just as the *New York Times* does not report "all the news that's fit to print," neither does any single service catch everything. At the urging of our readers, we have gone to great lengths to identify many "intermediates" in organic synthesis that are not indexed elsewhere. Consequently, many intermediates (i.e., unisolated reaction intermediates) indexed for *Index Chemicus Online* cannot be retrieved through any other chemical information service.

As with our other services, we expect that the heaviest users of *Index Chemicus Online* initially will be those who already subscribe to *CAC&IC*. While it will be possible to browse online, we expect that its main use will be for retrieval. For current awareness, *CAC&IC* along with *CC* would still be used weekly. This is not to say that *Index Chemicus Online* would never be used for browsing. The chemist does browse retrospectively. For example, if you want to explore the data base for classes of compounds that exhibit a particular biological activity, a search could be performed in the bibliographic file and compounds viewed on the structure file. In this way, the graphics capabilities of *Index Chemicus Online* are vital. You can explore the literature in creative ways not previously practical.

Although *Index Chemicus Online* is an interesting file that is fun to use, it was designed as a sophisticated search and display tool for chemists and information specialists, typically employed in pharmaceutical or chemical research. This applies in academia as well as in industry. We fully expect that educators will use *Index Chemicus Online* to

Table 1: Displayable fields in *Index Chemicus Online*™

AP—AUTHOR AFFILIATION
AU—AUTHOR
*BA—BIOLOGICAL ACTIVITY
CP—COMPOUND-RELATED DATA (SEQUENCE NUMBER AND COMPOUND NUMBER)
DT—DOCUMENT TYPE
*ED—LEVEL OF EXPERIMENTAL DETAIL
*EX—EXPLOSIVE REACTION ALERT
FC—MICROFICHE COORDINATES TO THE *CURRENT ABSTRACTS OF CHEMISTRY AND INDEX CHEMICUS*® (*CAC&IC*®) MICROFORM CUMULATION
FL—MICROFILM COORDINATES TO THE *CAC&IC* MICROFORM CUMULATION
*IM—INSTRUMENTAL METHOD
*ISO—ISOTOPE
*IT—INDEX TERM
*LA—LANGUAGE OF DOCUMENT
*MA—MAILING ADDRESS
MF—MOLECULAR FORMULA
NO—*CAC&IC* ABSTRACT NUMBER AND *THE GENUINE ARTICLE*™ (FORMERLY *OATS*®) ORDER NUMBER
*NSM—NEW SYNTHETIC METHOD ALERT
SO—SOURCE
TI—TITLE
WLN—WISWESSER LINE NOTATION

*These fields do not appear in every record.

demonstrate state-of-the-art methods of chemical information retrieval to students. Indeed, those students who are not at least familiar with such methods may find themselves at a disadvantage once they become working scientists.

As mentioned here earlier, Telesystemes is the vendor for *Index Chemicus Online*. Telesystemes is a private company which receives considerable support from the French government.[2] It is a major vendor and has mounted many other data bases. In fact, Telesystemes was the first vendor to provide graphic substructure access to the original *Chemical Abstracts* file. In short, they are experienced in operating an online system. Although Telesystemes's main computer is based in Valbonne, on the French Riviera, all US and European customers can reach it via a telephone call to their local Telenet, Tymnet, or other telecommunications hosts.

There are key software packages and command languages involved in using Telesystemes's files. The *Questel* software deals with bibliographic data. The *DARC* system, pioneered originally by Jacques Dubois, at the University of Paris,[3] deals with chemical structure data. We chose Telesystemes as a vendor because it included *DARC*'s superior graphics capabilities. Through *DARC*, you can view full-screen diagrams of the compounds you retrieve from a search.

You can also use diagrams of compounds to *search* with *DARC*. This can be done by using any of several graphic input techniques or by text input. Figure 1 shows a sample search query, while Figure 2 provides an example of a structure "hit" that was subsequently retrieved in response to the query.

You can search the bibliographic, or *Questel*, portion of the file by the traditional access modes including author, journal source, and keyword. Figure 3 shows a sample bibliographic hit for the query given in Figure 1. Additionally, the alerts to biological activities or new synthetic methods can be used as access modes. For example, a search for new methods for synthesizing dihydrobenzofuran retrieved the bibliographic record shown in Figure 4. A corresponding structure record from the *DARC* file is shown in Figure 5.

If you do not have a graphics terminal, you can still search for chemical information using text input, molecular formulae, or Wiswesser Line Notation

Figure 1: Sample structural search query. The search query is: Have any derivatives of 11H-dibenz(b,f)-1,4-oxathiepin been made?

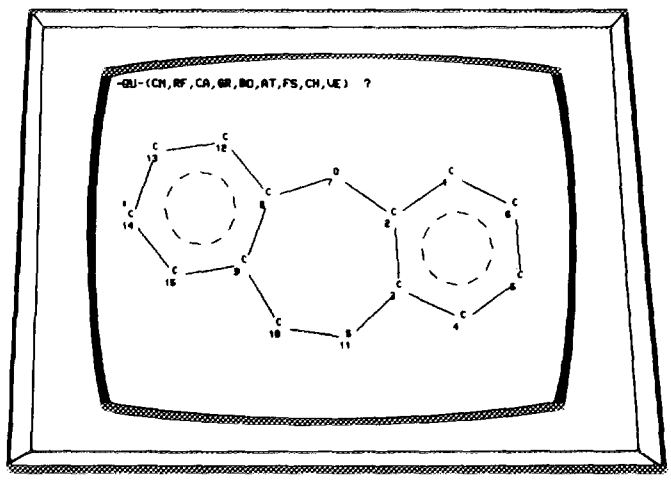


Figure 2: Display of first answer from substructure search on 11H-dibenz(b,f)-1,4-oxathiepin.

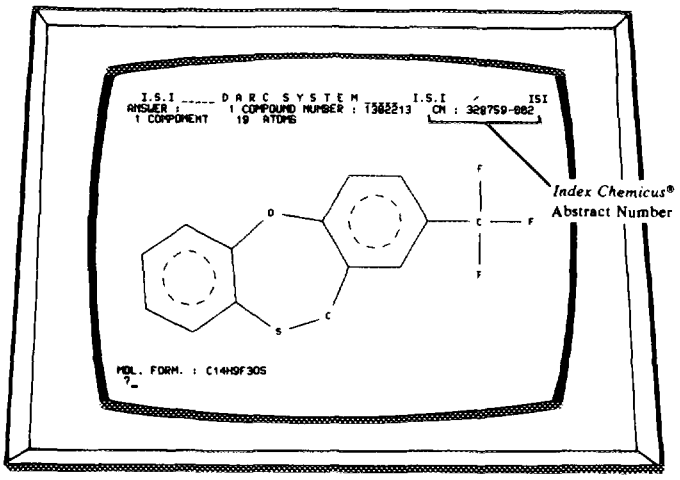


Figure 3: Bibliographic record corresponding to structural record shown in Figure 2.

Index Chemicus® Abstract Number

-1- 122861 C.ISI IC ONLINE
NO : 328759 OATS ORDER: NN659
TI : NEUROTROPIC AND PSYCHOTROPIC AGENTS. 164. TRICYCLIC PSYCHOTROPIC AGENTS CONTAINING TWO CHALCOGEN ATOMS IN THE CENTRAL RING: DERIVATIVES OF 11H-DIBENZ(B,F)-1,4-OXATHIEPIN.
AU : SINDELAR K.; HOLUBEK J.; RYSKA M.; DLABAC A.; METYSOVA J.; SCATEK E.; HRUBANTOVA M.; PROTIVA J.; PROTIVA M.
AF : RES INST PHARM & BIOCHEM, 130 60 PRAGUE 3, CZECHOSLOVAKIA.
SO : COLLECT CZECH CHEM COMMUN; V47(3); P.967-83; 1982
LA : ENG (ENGLISH)
DT : J (ARTICLE)
IT : — DIBENZOXATHIEPIN(B,F)(1,4): 11H, DERIVS, PSYCHOTROPIC AGENTS, SYN
BA : ANTIRESERPINE ACTIVITY; ANTIAPOMORPHINE ACTIVITY; ANTIDEPRESSANT ACTIVITY; ANTIBACTERIAL ACTIVITY; ANTIFUNGAL ACTIVITY
IM : THIN LAYER CHROMATOGRAPHY; INFRARED SPECTRA; NUCLEAR MAGNETIC RESONANCE; MASS SPECTRA; ULTRAVIOLET SPECTRA; COLUMN CHROMATOGRAPHY
CP : — 001(1B)
— MF : C13H9C10S
— WLN: T_C676_BO_IS_JHJ_MG
— 002(1C)
— MF : C14H9F3OS
— WLN: T_C676_BO_IS_JHJ_MXFFF
— 003(2A)
— MF : C14H10O3S
— WLN: T_C676_BO_IS_JHJ_JVQ
— 004(3A)
— MF : C16H15NO2S
— WLN: T_C676_BO_IS_JHJ_JVN1&1
— 005(4A)
— MF : C16H17NOS
— WLN: T_C676_BO_IS_JHJ_J1N1&1
— 006(5A)
— MF : C17H19NOS
— WLN: T_C676_BO_IS_JHJ_J2N1&1
— 007(6A)
— MF : C18N2INOS
— WLN: T_C676_BO_IS_JHJ_J3N1&1
— 008(6B)
— MF : C18H20C1NOS
— WLN: T_C676_80_IS_JHJ_J3N1&1_MG
FC : 00793-B13
FL : 2047M817G

Figure 4: Bibliographic record from a search for a new synthetic method and dihydrobenzofuran.

Accession Number

?. . SH MAX
-1- 106213 C.ISI IC ONLINE
NO : 312111 OATS ORDER: . . . . .
TI : THE SYNTHESES AND ABSOLUTE CONFIGURATIONS OF FOMANNOXIN, (—)-5-ACETYL-2-(1-HYDROXYMETHYLVINYL)-2,3-DIHYDROBENZOFURAN, AND ANODENDROIC ACID.
AU : KAWASE Y.; YAMAGUCHI S.; INOUE O.; SANNOMIY M.; KAWABE K.
AF : TOYAMA UNIV, FAC SCI, DEPT CHEM, TOYAMA 930, JAPAN.
SO : CHEM LETT; (12); P.1581-4; 1980
LA : ENG (ENGLISH)
DT : J (ARTICLE)
IT : — FOMANNOXIN: SYN & ABS CONFIG DETERMINATN
— BENZOFURAN: 5-AC-2-(1-OH-ME-VINYL)-2,3-DI-H, SYN & ABS CONFIG DETERMINATN
— ANODENDROIC ACID: SYN & ABS CONFIG DETERMINATN
IM : INFRARED SPECTRA
NSM : NEW SYNTHETIC METHOD
CP : — 001(7)
— MF : C13H1403
— WLN: T56_BOT&J_CYU1&1OV1
— 002(8)
— MF : C15H1604
— WLN: T56_BOT&J_CYU1&1OV1_GV1
— 003(9)
— MF : C13H1603
— WLN: T56_BOT&J_CX1&1&OV1
— 004(10)
— MF : C14H1604
— WLN: T56_BOT&J_CX1&1&OV1_GVH
FL : 2461M016G
FC : 00749-D11

Figure 5: Structural record corresponding to bibliographic record shown in Figure 4.

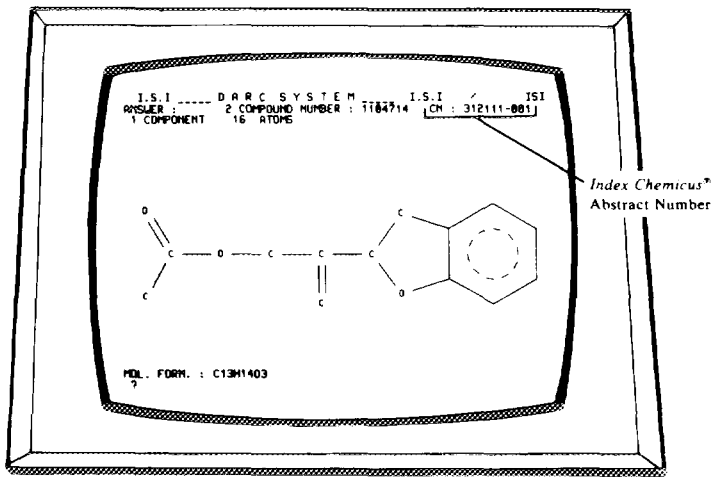


Figure 6: *CAC&IC*® record corresponding to sample *Index Chemicus Online*™ hits shown in Figures 2 and 3.

328759

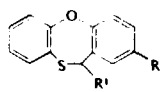
NEUROTROPIC AND PSYCHOTROPIC AGENTS. 164. TRICYCLIC PSYCHOTROPIC AGENTS CONTAINING TWO CHALCOGEN ATOMS IN THE CENTRAL RING: DERIVATIVES OF 11H-DIBENZ(B,F)—1,4-OXATHIEPIN.

SINDELAR K, HOLUBEK J, RYSKA M, DLABAC A, METYSOVA J, SVATEK E, HRUBANTOVA M, PROTIVA J, PROTIVA M.
RES INST PHARM & BIOCHEM, 130 60 PRAGUE 3, CZECHOSLOVAKIA.

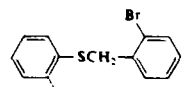
COLLECT CZECH CHEM COMMUN 47(3), 967-83(1982).

Reactions of 2-bromobenzyl bromide and its analogues *XVII* and *XXV* with 2-hydroxythiophenol resulted in 11*H*-dibenz[*b*,*f*]-1.4-oxathiepin (*Ia*) and its 2-chloro (*Ib*) and 2-trifluoromethyl derivative (*Ic*). Treatment of the lithium compounds derived from *Ia* and *Ib* with carbon dioxide and dimethylaminoalkyl chlorides gave compounds *IIa*, *Va* and *VIab*; modification of the side chains led to amines *IVa*, *VIIa* and *VIIIa*. 11-(1-Methyl-4-piperidyl) derivatives *Xbc* were obtained by chlorination of compounds *Ibc* with sulfuryl chloride or N-chlorosuccinimide and the following treatment with 1-methyl-4-piperidylmagnesium chloride. Compound *Ib* was transformed by oxidation to the sulfone *XX* affording by treatment with sodium hydride and tert-aminoalkyl chlorides the basic sulfones *XXI* and *XXII*. While the nuclearly unsubstituted amines with the aliphatic side chains (*IVa* and *VIIa*) have intensive antireserpine activity and are potential antidepressants, the 11-(1-methyl-4-piperidyl) derivatives with a substituent in position 2 of the skeleton (*Xbc*) are potential neuroleptics; the trifluoromethyl derivative *Xc* especially has outstanding cataleptic and antiapomorphine efficacy.

ANTIRESERPINE ACTIVITY
ANTIAPOMORPHINE ACTIVITY
ANTIDEPRESSANT ACTIVITY
ANTIFUNGAL ACTIVITY
ANTIBACTERIAL ACTIVITY



I(a,b,c),IIa-Va,VI(a,b), VIIa,VIIIa,IX(b,c), X(b,c),XIb,XXVIa



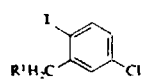
XII,XIII

R¹
XII: Me
XIII: H

	R	R¹
Ia:	H	H
b:	Cl	H
c:	CF_3	H
IIa:	H	COOH
IIIa:	H	$CONMe_2$
IVa:	H	CH_2NMe_2

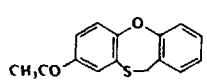
	R	R¹
Va:	H	$(CH_2)_2NMe_2$
VIa:	H	$(CH_2)_3NMe_2$
b:	Cl	$(CH_2)_3NMe_2$
VIIa:	H	$(CH_2)_2NHMe$
VIIIa:	H	$(CH_2)_3NHMe$

	R	R¹
IXb:	Cl	Cl
c:	CF_3	Cl
Xb:	Cl	1-Me-4-piperidyl
c:	CF_3	1-Me-4-piperidyl
XIb:	Cl	1-Me-4-HO-4-piperidyl
XXVIa:	H	COCl

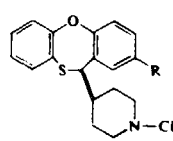


XVI-XVIII

R¹
XVI: OH
XVII: Br
XVIII: 2-HOC_6H_4S

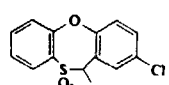


XV

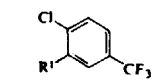


XIX(a,b,c)

	R
XIXa:	H
b:	Cl
c:	CF_3



XX-XXII



XXIII-XXV,XXVII

R¹
XX: H
XXI: $(CH_2)_2NMe_2$
XXII: $(CH_2)_3$piperidino
XXIII: CHO
XXIV: CH_2OH
XXV: CH_2Cl
XXVII: NN^+ Cl^-

(WLN) in the bibliographic file. As described on many earlier occasions, ISI has pioneered the use of this notation. WLN permits you to do many types of substructure or parent compound searches. We have encoded over 3,000,000 compounds to date and continue to input structural information to the data base in this manner. In order to provide graphic display of these compounds through the *DARC* system, we have developed programs to convert these WLNs to a format compatible with the *DARC* software. Therefore, we can encode private files as well, and convert them to formats compatible with *DARC* and other in-house graphics software packages.

The *Index Chemicus Online* data base was designed to be used in concert with other ISI services. For example, our new *CCI* will be cross-referenced with *Index Chemicus Online*. The *CCI*, which will be a separate online file, will include entries for all articles abstracted in *CAC&IC*. It will provide full coverage of about 300 core chemistry journals, and selective coverage from the rest of the *SCI* file. Each article will be tagged with its *IC* abstract number when appropriate. For example, if you have done a search in *CCI* and found a key relevant article, you can use the *IC* abstract number to retrieve and examine the structural diagrams for the compounds reported. Additionally, you can perform a cited reference search of articles retrieved in *Index Chemicus Online*. The *CCI* file will also have the research front capability that will add an interesting new dimension to the retrieval of chemical compounds. It will be a modern complement to *Beilstein*.

The *CAC&IC* 22-year Microform Cumulation will also be cross-referenced with *Index Chemicus Online*. By using the microform coordinate numbers which are included in each *Index Chemicus Online* record (see Table 1), you can go to the cumulation to retrieve the author's abstract and a hard copy of the structures shown. This feature would be particularly useful if your initial system does not have graphic capabilities. The cumulation, incidentally, covers the entire *CAC&IC* printed file from 1960 to 1981. (Annual cumulations are available for 1982 on.) Available in both microfiche and microfilm, it provides archival quality copy. Its price is $10,000 for industrial users, and $5,000 for educational institutions.

Figure 6 shows the *CAC&IC* record corresponding to the query given in Figure 1. For the time being, there is only one major difference between the coverage of *CAC&IC* and *Index Chemicus Online*. *CAC&IC* includes the author's abstract when available, while *Index Chemicus Online* does not.

To use an online system, you ordinarily need to learn a new command language. We have designed the *Sci-Mate*™ software package to overcome this requirement.[4] *Sci-Mate* is a microcomputer-based software system for maintaining private files and interacting directly with large online data bases. By the end of this year, we expect to introduce a new version of *Sci-Mate* that will include *Questel* as one of the many "host" options to search bibliographically.

We believe that the cost of *Index Chemicus Online* is quite reasonable. Bibliographic searching costs $70 per hour. Structure searching costs $160 per hour. The average search is approximately six to ten minutes and costs about $20. In addition, there is a charge of $.16 per record display.

There is no subscription fee, monthly minimum, or initiation charge. All you need to get started is a Telesystemes password. In the US, phone 800-424-9600, or write Questel Inc., 1625 I Street, Washington, DC 20006. In Europe, phone 33-1-544 38 13, or write Telesystemes, 40 Rue du Cherche Midi, 75006 Paris, France. Specify that you

want access to *Index Chemicus Online*. After you complete and return the user agreement you will receive your account number, password(s), and a complete set of *DARC/Questel* manuals, usually within one week.

A do-it-yourself kit for *Index Chemicus Online* will soon be available. It will include a primer which introduces the data base and a step-by-step tutorial manual for programmed instruction. The manual is designed to be used while you are sitting at the terminal. Half of each page shows what is on the screen, while the other half provides notes and instructions.

ISI will offer training sessions in major cities in the US and Europe. Whenever possible, these sessions will run concurrently with Telesystemes's training workshops.

So far, we have demonstrated *Index Chemicus Online* at several major conventions. It was first introduced at the International Online meeting held in London in December 1983. It was then demonstrated at the April National Online Meeting, in New York City, and the spring meeting of the American Chemical Society (ACS), in St. Louis, Missouri. In each case, *Index Chemicus Online* was very well received. Many chemists learned of ISI's long-term involvement in chemical information for the first time. The next meeting of the ACS will be held from August 26 to 31 in Philadelphia, Pennsylvania. You can be sure ISI and *Index Chemicus Online* will be there.

We have established a special hotline strictly for *Index Chemicus Online*-related inquiries. In the US, the number is 800-523-1857. Readers outside of the US should call 215-386-0100, ext. 1291. *Index Chemicus Online* subscribers will also receive Telesystemes's *DARC* and *Questel* newsletters, *Tel-a-Chem* and *Questel-a-Gram*. These newsletters will provide a user's forum, and feature sample searches, announcements, pricing information, and the like.

We have designed *Index Chemicus Online* to make it possible to use organic chemical information as part of the dynamic process of designing new molecular entities in medical, agricultural, and other applications. Whether you are a pharmaceutical chemist, microbiologist, or organic chemist you will want to test out this new methodology. Simply call our hotline, or write to Keri Luiso, Chemical Information Division Marketing, ISI, 3501 Market Street, Philadelphia, PA 19104.

* * * * *

My thanks to Brad Schepp for his help in the preparation of this essay.

REFERENCES

1. **Garfield E.** Introducing *Current Chemical Reactions. Essays of an information scientist.* Philadelphia: ISI Press, 1981. Vol. 4. p. 12-5.
2. **Bonnet J C.** *Questel/DARC* as an important service to the chemical industry. *Inform. Chim.* 231:243-5, 1982.
3. **Dubois J E.** French national policy for chemical information and the DARC system as a potential tool of this policy. *J. Chem. Doc.* 13:8-13, 1973.
4. **Garfield E.** Introducing *Sci-Mate*—a menu-driven microcomputer software package for online and offline information retrieval. Part 1. The *Sci-Mate Personal Data Manager*. Part 2. The *Sci-Mate Universal Online Searcher*. *Current Contents* (12):5-12, 21 March 1983 and (14):5-15, 4 April 1983.