

# Current Comments

## ISI's On-Line System Makes Searching So Easy Even a Scientist Can Do It: Introducing METADEX Automatic Indexing & ISI/BIOMED SEARCH

Number 4

January 26, 1981

Searching data bases on-line has certainly come of age. Several million on-line searches are performed each year. These searches require the use of a computer keyboard or terminal that is tied into a computer, usually located thousands of miles away. For example, your library can use MEDLINE to access *Index Medicus*.<sup>1</sup> Or it may use Lockheed's *Dialog* system to access *Excerpta Medica*,<sup>2</sup> *Biological Abstracts*, and many other indexes. SDC, Bibliographic Retrieval Services (BRS), and other systems also provide on-line access to a wide variety of data bases. For example, ISI's *SCISEARCH*,<sup>3</sup> the on-line version of the *Science Citation Index*® (*SCI*®), is now available through Lockheed and will be available through BRS in the near future. *Social SCISEARCH*®, the on-line version of the *Social Sciences Citation Index*® (*SSCI*™), is available through SDC's *Orbit*, BRS, and Lockheed.

While the various on-line languages such as *Dialog*, *Orbit*, or BRS's modified *Stairs* have served a useful role to date, they leave much to be desired. They are not completely adequate and efficient even for conventional systems. So from the earliest days of on-line searching we realized that none of the existing systems was adequate to meet the needs of *SCI* and *SSCI* users. These systems were not designed with citation searching in mind. Considering this, it is remarkable how often our data bases are accessed.

One "problem" with the existing systems is that they are not "transparent" to the uninitiated user. You can take special courses to learn how to use each system. I tried one and found it rather uninspiring. But the most important factor is that you don't remember many of the details unless you are searching every day. That is why most on-line searches are done by librarians or information specialists whose main responsibility is on-line searching. When I do use the terminal in my apartment I follow a precise protocol I've developed and limit my searches to those which fall within this limited set of parameters.

A second problem with *SCISEARCH*, not apparent to the user, is its size. It requires more characters of computer storage than most other data bases. We must not only store standard bibliographic information for each current article we index, but we also store the references cited in each article. And since the *SCI* covers over 500,000 articles and book chapters each year, about one billion characters of storage are required.

To resolve these problems and others, we decided that a new software approach was needed as well as another approach to file organization. This highly transparent system is now being tested at several institutions in North America and Europe. The file covers only the biomedical portion of the *SCISEARCH* data base—nearly 1,300 of the most significant journals. For this

reason it is named ISI/BIOMED SEARCH™. The journals were selected by citation analysis and specialist judgment. They represent the core literature of biomedicine. This is considerably less than *SCI's* 3,000 or so journals. And to further reduce the size of the file, we have limited it to only the literature published from 1979 to the present.

Apart from simpler protocols for on-line searching there is another significant difference in the ISI/BIOMED SEARCH system. It is so transparent that even you and I can use it because of an important breakthrough in indexing methodology. This system of automatic indexing, based primarily on co-citation analysis which I have described recently,<sup>4,5</sup> provides a whole new dimension to literature searching because it is user-oriented. It is, in fact, directly derived from the natural and citation language of scientists themselves. For lack of a better term we call this new search technique "research front searching." But before going into details about the way articles are assigned to research fronts, let me show you how easy it will be to use ISI/BIOMED SEARCH.

Suppose you are interested in interferon. You can of course (as in the other systems I've mentioned) key in that term as a title word. You will find that from 1979 to date, there have been about 1,000 papers with interferon in their titles—much too large a bibliography for most purposes. In most systems, you would have to modify your search on interferon by adding other qualifying terms.

But with ISI/BIOMED SEARCH you can turn to a printed index of research front specialty names and see the names of current research specialties on aspects of interferon research (Figure 1). The index covers over 1,500 "hot" research areas in biomedicine which may be searched by the new technique of research front searching. And this list will be expanded regularly as we gain experience with the system.

If you find a research front name that reflects your interest, you key in the

Figure 1: Entry from 1980 *Research Fronts in ISI/BIOMED SEARCH™*, showing the names of all research fronts which use "interferon" in their titles. Beside each name is its serial number which may quickly be keyed into the computer to access the research front.

<b>INTERFERON</b>	
<b>INTERFERON and the IMMUNE-SYSTEM</b>	1980-1474
<b>INTERFERON as an ANTIVIRAL agent</b>	1980-1145
<b>INTERFERON production, PURIFICATION and characterization</b>	1980-0639
<b>MECHANISMS of action of INTERFERON</b>	1980-0516
<b>PHARMACOLOGY of IMMUNE INTERFERON</b>	1980-1543
<b>PROTEIN synthesis INITIATION and regulation by INTERFERON and PROTEIN-KINASE</b>	1980-0036
<b>Production and PURIFICATION of human INTERFERON</b>	1980-0485

serial number which appears beside the name in the index. This number system saves you the time you would have spent typing in the name itself.

So if your area of interest is reflected in the research front specialty called *Protein Synthesis Initiation & Regulation by Interferon & Protein-Kinase*, you simply key in 1980-0036 and find there are 168 recent papers published on that topic or research front.

If the bibliography is too lengthy for your needs, you may want to further refine it before you begin to print out the items. This may be done in a number of ways. You may request only those articles most relevant to the specialty, that is, those articles which cite the most articles in the specialty. (More about these hidden citation links later.) You may narrow the bibliography by limiting it by type of document (e.g., reviews, meeting abstracts, etc.) or language, or you may request only those documents which appear in two specialties. You may also limit by year or accession number to get only those items added to the data base during a particular year or since you last searched the data base. This last capability permits you to use ISI/BIOMED SEARCH for current awareness.

Once you are satisfied with the size of the bibliography, you can print out the bibliographic descriptions of these articles. Figure 2 gives a sample of the information you will receive. Note the last

line in the entry, which begins "Explore Specialty...." There follow the serial numbers of the specialties to which that paper has been assigned, thus providing information to move on to another specialty.

When you key in one of the serial numbers for another specialty to which the article belongs, 1980-0643, for example, the computer will display its name, *Terminal-Cap-Structure in Viral Messenger-RNA*, and the number of papers indexed under that research front. You can then decide if the research front is interesting enough to warrant further searching.

The indexing methodology behind the new file rests on the fact that the intellectual structure of biomedical research can be defined as a network of research front specialties: discrete, but interacting, and sometimes overlapping areas of intense research activity. The pattern of interaction between them comprises a map of the cutting edge of biomedical research.<sup>6</sup>

Research front specialties are defined through a process, mentioned earlier, called "co-citation clustering." Each cluster is a group of highly cited, closely related papers—in effect, the core literature of a given specialty. Each cluster represents the most significant work in a particular research area.

Our first step in assembling these clusters is to identify articles cited a

given number of times—say, 15—by the source articles in the *SCI*. This gives us a group of about 25,000 articles which are among those considered most important by publishing researchers. Only a fraction of these are the *Citation Classics* discussed each week in *Current Contents*<sup>®</sup>. Our next step is to group together the articles that are cited together, or co-cited, by the source publications. This yields a smaller group that are linked to one another by the most recent literature. Many of these appear in our various studies of most-cited papers.

In citation analysis, we assume that highly cited articles represent the significant active concepts or methods in a field. On the basis of our research, we've also found that co-citation relationships between papers identify highly specific concepts or methods.

Since our cluster programs are run at least yearly, the articles assigned to each cluster are those highly cited in the most recent literature. Unlike traditional *a priori* classification systems, the ISI/BIOMED SEARCH system is *a posteriori*.<sup>7</sup>

For lack of a better term, I would provisionally choose to call the indexing language ISI AUTOCLASS or METADEX. In linguistics we refer to metalanguage, the language used outside the text. By using the authors' citations to other works we are silently using the

**Figure 2:** A typical item from a printout of items retrieved by searching the research front *Protein Synthesis Initiation & Regulation by Interferon & Protein-Kinase*.

Author	Article Title	Document Type
Language	*INTRACELLULAR EVENTS IN INTERFERON-TREATED CELLS (ENGLISH/BIBLIOGRAPHY, REVIEW)	
	HOVANESSIAN AG	Author's Address
	INST PASTEUR, DEPT VIROL, UNITE ONCOL VIRALE, 25 RUE DU DR ROUX, F-75724 PARIS, FRANCE.	
	DIFFERENTIATION 15(N3):139-151 1979	154 REFERENCES
	ICI ACCESSION #80018775; ISI OATS ORDER #JC437	
	EXPLORE SPECIALTY 1980-0036; 1980-0516; 1980-0020; 1980-0643; 1980-1145	
Journal Title	Volume, Issue, Pages, and Year	No. of references in article
		Research fronts to which this article belongs

language "meta" to the document being indexed. Citation language is the meta-language of science or scholarship.<sup>8</sup> (p. 179)

To facilitate this process, ISI subject specialists first study the titles of papers in each cluster. They create specialty names by using the current terminology most often used in the citing papers. Thus, cluster names reflect the way researchers themselves are thinking and talking about these topics.

The METADEX methodology used in ISI/BIOMED SEARCH gives you (who may know or care little about indexing philosophy) three important benefits: 1. It allows you to take advantage of citation indexing without requiring that you begin with a particular cited author or paper. 2. It permits you to begin with a research front specialty name whose terminology is current and familiar. 3. It allows you to access data without a lot of special training or constant practice.

We believe the system exemplified by ISI/BIOMED SEARCH is a significant breakthrough in automatic indexing and on-line searching. It is our expectation that companion search facilities in the physical sciences, mathematics, social sciences, and humanities will be available as time, facilities, and energy permit. For each field mentioned, appropriate specialties will be identified and named. In fields like mathematics the citation and chronological parameters

will be different but the procedures quite similar. The significance of AUTOCCLASS or METADEX in such literature will be appreciated by anyone who has had to deal with the indexing of vague or abstract mathematical concepts expressed in mathematical equations.

I believe it is fitting to mention the obvious in closing. The system will of course permit you to do the most basic search in a citation index, that is, to find out who has cited a particular paper, book, journal, or author. You can also retrieve all the papers from your institution, country, etc., and you will be able to search all of those in combination with the kinds of specialty data already described.

I think your problem with ISI/BIO-MED SEARCH will be not how quickly you can find what you want but resisting the temptation to go on. While I continue to point out the significant advantages of printed indexes, I truly believe that these new systems will open up an entirely new approach and attitude towards using the literature. When this is combined with instant retrieval of papers we can truly visualize a significant transformation in the ways in which research will be conducted in the future. Scientists will be able to conduct research not only in the library or the lab but at home or anywhere else they wish to take a portable terminal.

©1981 ISI

## REFERENCES

1. **Garfield E.** Some reflections on *Index Medicus*. *Current Contents* (51):5-11, 17 December 1979.
2. .... *Excerpta Medica*—abstracting the biomedical literature for the medical specialist. *Current Contents* (28):5-10, 14 July 1980.
3. .... ISI's SCISEARCH™ time-shared system trades time for money—but are you ready for this? *Current Contents* (40):5-6, 4 October 1972.\*
4. .... ABCs of cluster mapping. Part 1. Most active fields in the life sciences in 1978. *Current Contents* (40):5-12, 6 October 1980.
5. .... ABCs of cluster mapping. Part 2. Most active fields in the physical sciences in 1978. *Current Contents* (41):5-12, 13 October 1980.
6. .... ISI's Atlas of Science may help students in choice of career in science. *Current Contents* (29):5-8, 21 July 1975.\*
7. .... Technology assessment, citation analysis, and invisible radiations. *Current Contents* (24):5-6, 14 June 1972.\*
8. **Morris C.** *Signs, language, and behavior*. New York: Braziller, 1955. 365 p.

\*Reprinted in: **Garfield E.** *Essays of an information scientist*. Philadelphia: ISI Press, 1980. 3 vols.