

When Browsers Become Retrievers,
CC's Natural-Language WSI's Help

September 22, 1975

Number 38

In this and next week's issue of *Current Contents*[®], all readers will find a *Weekly Subject Index (WSI)*. I hope those readers who don't now subscribe to *WSI* will take advantage of this opportunity to discover what other *CC*[®] subscribers already know: that the *Weekly Subject Index* is a valuable adjunct to scanning the contents pages of *CC*. Our surveys¹ have shown that most readers use *WSI* to complement scanning—to insure that no important articles have been overlooked. And *WSI* is also an excellent aid for quickly finding relevant articles *before* browsing through *CC*—as well as for relocating articles afterwards.

The *Weekly Subject Index* is a fairly new feature of *CC*. *WSI* first appeared as an *optional* feature of *CC/Life Sciences* in 1972. Thousands of readers quickly subscribed. This encouraged us to expand our *WSI* program to other editions of *CC*. In 1973, we added one to *CC/Physical & Chemical Sciences* on an optional basis, but included *WSI* as part of the then new *CC/Clinical Practice*. In 1974, we added *WSI* to *CC/Social & Behavioral Sciences*, and this year we added *WSI* to *CC/Engineering Technology & Applied Sciences* and *CC/Agriculture, Biology & Environmental Sciences*.

Today *WSI* is still available as an option to readers of *CC/Life Sciences* and *CC/Physical & Chemical Sciences*. It is included in the other four editions of *Current Contents*.

You may wonder, as I have, why a subject index to a browsing device like *CC* should be so popular. Taking a theoretical approach one could say it is useful redundancy. We know from communication theory that redundancy is necessary to eliminate errors in transmission of signals. The human eye is not perfect and under varying degrees of time or other pressure an important article may miss our attention. Checking a list of important terms may provide the insurance needed to avoid such misses. In addition, there is the problem of context. When a word appears in an index, it may well attract our attention in a way that it otherwise might not. We also know that few readers scan every contents page. Therefore, the subject or word index may turn up articles in these journals. This is especially true of large multidisciplinary journals.

So, if you recognize the value of the extra index, then you should examine the particular virtues of *WSI*. Since it is bound right into *CC*, it's always convenient. *WSI* is also quite current. *WSI* is just as timely as *Current Contents*.

We believe the *Weekly Subject Index* has been so successful because it's easy to use. *WSI* is based on a 'natural-language' system which some claim to be "optimally the best indexing language" for information retrieval in the sciences.² This assertion is subject to debate among indexing experts, but we believe it is the most cost-effective word retrieval method now available.

A natural-language subject index differs in several respects from a more conventional subject index in which professional indexers classify articles under an authorized subject-heading vocabulary. Problems sometimes occur because the subject-headings are slow to change and often do not adequately specify an emerging area of interest—and indexers sometimes make mistakes in classification. Furthermore, a user who does not know the idiosyncracies of the indexing vocabulary can waste time just trying to figure out under which term articles of interest are listed.

In *W/SI*'s natural-language system, the indexing terms are taken from the article titles themselves. In *CC* every contents page, including those in *Current Book Contents*[™] is indexed. (We expect to include titles from the *ISI Press Digest* very soon.) In effect, the authors themselves have indexed their articles by the title-words they chose.

CC readers who are active professionals can quickly find relevant articles by employing the terms actually in current use in their fields. For instance, in an index like *Index Medicus*, CELLULASE is not yet a subject heading. In *W/SI*, it would have been a heading the first time it appeared in a title.

Not every title-word becomes an indexing term in *W/SI*. Words without indexing significance ("the", "and", "is", etc.) are eliminated. In addition, some title words occur so frequently that it is more efficient to use them in combination with other terms. Thus 'word-phrases' or 'hyphenated terms' of two to four words are used. For example, IMMUNE-RESPONSE, QUARK-PAIR-CREATION, DNA-SYNTHESIS, BIRTH-CONTROL, etc. are among

thousands of such terms we edit by a human-machine system.

A guide to *W/SI* use appears on page S-2 of this issue. The *Weekly Subject Index* has incorporated a minimum of standardization. For example, variant spellings of the same word (behavior, behaviour) are unified. On the other hand, derivative forms of the same word (ion, ions, ionization) appear as separate indexing terms. Synonyms (such as DOWN'S-SYNDROME and MONGOLISM) also appear as separate headings.

Occasionally a user will locate a term in *W/SI*, turn to the contents page, but will not find the term in the indicated title. This is because title words in *W/SI* are taken from the title as they appear at the head of the article itself, not from the table of contents. Since contents pages occasionally use abbreviated titles, the term which appeared in *W/SI* may be missing from the contents page.

All titles in *CC* are translated into English before they are indexed in *W/SI*. Thus you may start with an English term in *W/SI* and be directed to an article written in Russian, French, German, etc. We always indicate the languages used on the contents page.

Each entry in *W/SI* indicates the *CC* page on which the table of contents begins. Since the contents of a journal issue may run to two or more pages in *CC*, the article may not appear on the first page indicated, but on a following page.

No index is perfect—or perfectly clear at first glance—but we have tried to make *W/SI* as simple and convenient as possible. Its increasing use by *CC* readers indicates that we are succeeding. Next week I'll tell you how we work at improving all of ISI's indexes.

1. Harris R M & Garfield E. Weekly subject index to *Current Contents*, a new aid to complement scanning scientific literature. Paper presented at the 1972 Fall Meeting of the American Chemical Society, New York City, 28-31 August 1972.
2. Bhattacharyya K. Effectiveness of natural language in science indexing and retrieval. *Journal of Documentation* 30(3):235-53, 1974.