

“Word-phrases” in CC’s
Weekly Subject Index (WSI) Increase
 Search Efficiency

May 3, 1972

Recently, ISI® concluded a small survey of CC® subscribers who receive the *Weekly Subject Index*¹⁻⁴ to *Current Contents®/Life Sciences*. Such surveys keep us informed of users’ needs and reactions to our services. Since the *Weekly Subject Index* began 3 months ago we have been aware of the need to improve our computer indexing methods to include “word-phrases”—a redundant expression to say the least.

In recent issues of the *WSI* we began to use “word-phrases”, otherwise known in the information industry as “hyphenated terms without hyphens”, “bound terms”, “coordinate terms”, etc. These word-phrases will speed up location of relevant articles. For example, articles on *cyclic AMP* can be found listed under that “word phrase” rather than under the separate terms *cyclic* and *AMP*. Like *cyclic AMP*, many of the newly used word-phrases are seemingly “obvious”. Consider *blood pressure*, *magnetic field*, *surface-active*, *x-ray diffraction*, etc. But what is obvious to one person is not to another. In my experience it is generally obvious only if one knows something about frequency of occurrence.

In the survey, and in correspondence with readers, I’ve learned that there is some confusion as to how the *WSI* is made. Some readers have criticized our “choice” of “subject-headings” or the

“grouping” of subjects under “terms” that are “too broad”, “too specific”, etc. Such criticism comes from a misunderstanding of how the index is put together. It is *not* made by selecting subject-headings or “descriptors” from a subject-heading list or thesaurus. It would be impractical to do “cerebral” indexing in the very short time available. We have about eight hours between the close-out of any weekly issue of *Current Contents* and dispatch of the completed copy to the printer. Indexing is performed by computer selection of *every* significant word from *every* article title in the issue. The computer also arranges these title words alphabetically, while keeping track of the articles and journals from which they came and the pages of CC on which they will appear. Therefore, the index terms are words used by the authors; they are not “selected” from some static list. The “vocabulary” of *WSI* is as alive and as current as the vocabulary of the scientist-writers or or editors responsible for selecting the article titles.

For this variety and flexibility a certain price is extracted in terms of precision, but countless studies have shown that title-word indexing is no less cost-effective⁵ than other methods.

Just as we have not intuitively selected headings for individual articles, so we haven’t intuitively selected word

phrases we thought might be most “useful”. *Blood-pressure* would have been an obvious choice, if we had set about making choices. There may be some researcher who’s interested in the concept of *pressure* wherever it manifests itself—in the cardiovascular system, in hydraulic jacks, in vacuum pumps, etc.—who would have preferred we leave *pressure* alone. I’ll probably hear from him soon. Others will be glad to find that *blood pressure* is separated from all other pressures of life.

To others, *lung cancer* might have seemed equally obvious. Not so. In the “obvious” case of lung cancer, there are many researchers who still want to follow all kinds of cancers, whatever their etiologic, pathologic, anatomic tags.

For *WSI* we’ve been able to adopt a frequency-based methodology because we have such data. We process some 400,000 article titles each year for the *Permuterm*[®] *Subject Index*⁶ of the

Science Citation Index[®]. For more than six years we’ve kept track of word-pair frequencies in *PSI*. This information is used to determine whether or not a particular word-phrase appears in the *WSI*.

I hope that readers who have already come to depend on the *WSI* will find that the addition of word-phrases increases its value both for weekly scanning and retrospective retrieval. Readers who haven’t gotten around to testing the *WSI* now have one more reason to do so. Many *WSI* users have told us that it saves substantial time each week, but it especially adds to one’s peace of mind to have this added insurance while scanning. The real payoff comes a few weeks later when you are trying to remember just exactly where you saw that article by Blogs or Claghorn. Imagine how impressive it is during the coffee-klatch or lunch-time seminar to pull out *WSI* and nail the reference down in seconds.

1. Garfield, E. ISI announces availability of a *Weekly Subject Index* to *Current Contents/Life Sciences*. *Current Contents* No. 34, p. 5-6, August 25, 1971.
2. ----- . The *Weekly Subject Index* makes for zero delay in current awareness and information retrieval. *Current Contents* No. 37, p. 4-5, September 15, 1971.
3. ----- . A *Current Contents* Odyssey from 200 to 1100 journals plus a *Weekly Subject Index (WSI)* with zero time-lag. *Current Contents* No. 1, p. 5-6, January 5, 1972.
4. ----- . A *Weekly Subject Index* for discovery and recovery of information in the life sciences. *Current Contents* No. 3, p. 5-6, January 19, 1972.
5. Cleverdon, C. Cranfield tests on index language devices. *ASLIB Proc.* 19(6):173, 1967.
6. Garfield, E. *Permuterm Subject Index*, the primordial dictionary of science. *Current Contents* No. 22, p. 4, June 3, 1969.