LETTERS

Retrieval of scientific information

I was very much interested in the article entitled "All the Red-Legged Partridges" (PHYSICS TODAY, Nov. 1965, p. 57) by Katherine Way, director of the Nuclear Data Project at Oak Ridge National Laboratory. Dr. Way made a strong case for retrieving an extremely high percentage of necessary data.

However, her proposal to have authors attach index words or key words to their own papers will not accomplish the high degree of accurate retrieval that she describes. At IBM we have a system of indexing-at-source, guided by an IBM list or thesaurus, and we have not been satisfied with retrieval results attained by the matching of labels. Therefore I take issue with Dr. Way's expectation that 99.5% retrieval would be accomplished in this manner.

I believe that normal text searching with computers, as used in the IBM technical information retrieval center, would come closer to achieving the goal. Good question-phrasing to a computer, however, would depend on the retrieval-specialist's knowledge of the specific data base concerned. Hence the achievement of 99.5% retrieval is something that we can not promise our customers. There are occasions when it may be achieved, but as a general occurrence it would be rare indeed.

The improbability of a human reader's achieving the 99.5% goal should be a warning since computers, though fast and accurate, can not make necessary interpretations. The computer is an extremely useful tool for information retrieval, but it, too, has limitations. It matches terms mechanically, using such Boolean-like logic as OR, AND, BUT NOT, YES, ADJACENT WORDS, etc., but it can not think or judge answers as a man can while reading.

I believe our combination of normal text-searching and the use of authors' key words approaches the retrieval satisfaction desired by Katherine Way. But without specific experiments on the particular data base used, I do not believe that 99.5% retrieval can be accomplished as an operating procedure.

Joseph J. Magnino Jr.
Manager, IBM Technical Information Retrieval Center

Author's comment

A phone call to Mr. Magnino's office brought out the information that the unsatisfactory results in author-indexing experiments arose in connection with the use of a 20,000-word thesaurus. To me this does not seem very surprising. Most people unfamiliar with indexing would surely feel somewhat overwhelmed if confronted with such a compendium. However, I have learned that the Bureau of Reclamation of the US Department of the Interior reports "satisfactory" results with author key-wording by means of a fairly large thesaurus, one having some 3,000 terms. It is not yet clear what "satisfactory" means to the Bureau of Reclamation or "unsatisfactory" to IBM. The quantitative results may be the same in both cases.

The suggestion in my article was for short thesauri in the different fields of science, thesauri that could be written, as in my example, in the natural words or symbols of a field. There is no comprehensive evidence yet as to how well authors would perform the indexing or key-word-assigning task with the aid of such thesauri. Preliminary results with our key-word list have been good. However, carefully controlled experiments would be very worthwhile. Perhaps there is a limiting size to the thesaurus with which an inexperienced indexer could cope.

To my mind the important thing for scientists is that they begin to concern themselves with designation of useful key-word or indexing terms in their own fields, because indexing terms are the foundation blocks of any storage and retrieval system. Many searching systems will certainly be developed, but none will be able to locate words or concepts not built in.

Katharine Way
Oak Ridge National Laboratory