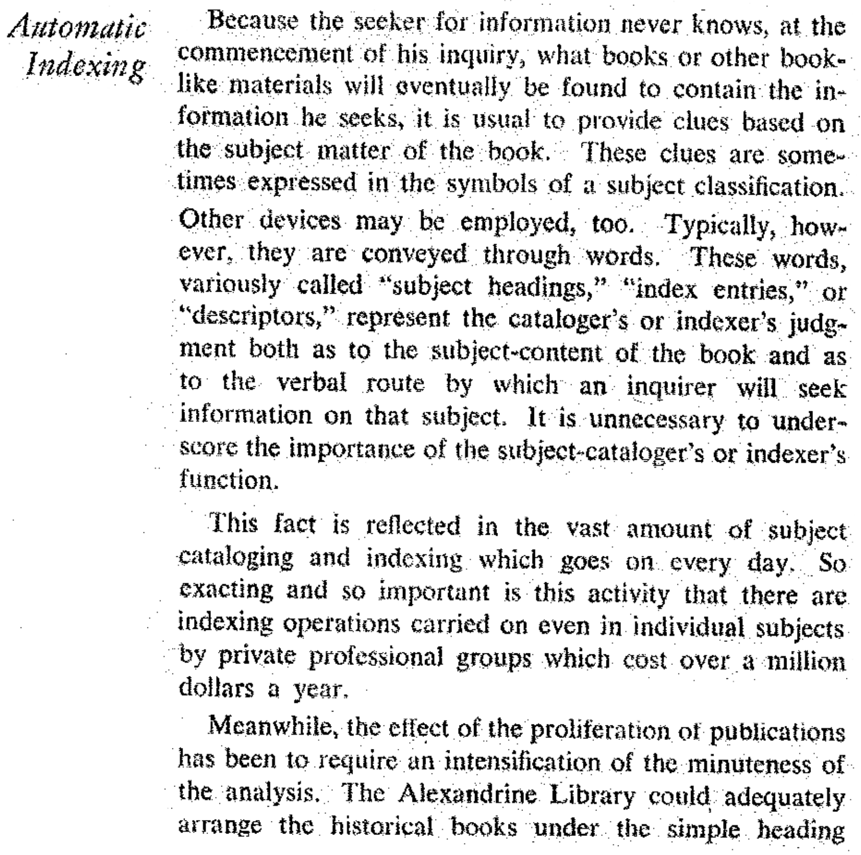
1959: "Automatic Indexing" was one of nearly 40 grants CLR awarded during its third year of operation. The project was granted $68,216.

Project PI Don R. Swanson was an information scientist most known for his work in literature-based discovery in the biomedical domain. In the 1980s, he wrote extensively on the concept of "undiscovered public knowledge," which exists as independently created fragments that are logically related but never retrieved, brought together, and interpreted.

The CLR grant sought to explore whether machines could be devised to recognize the meaning of words and apply the correct subject entry to reflect that meaning. The result of Swanson's work was reported in "Searching Natural Language Text by Computer," *Science* 132: 1099–1104, Oct. 21, 1960. His findings were sufficiently promising that CLR awarded a subsequent grant.

The following description was provided in the third annual report of the Council on Library Resources, Inc., for the period ending June 30, 1959, pp. 26-28.

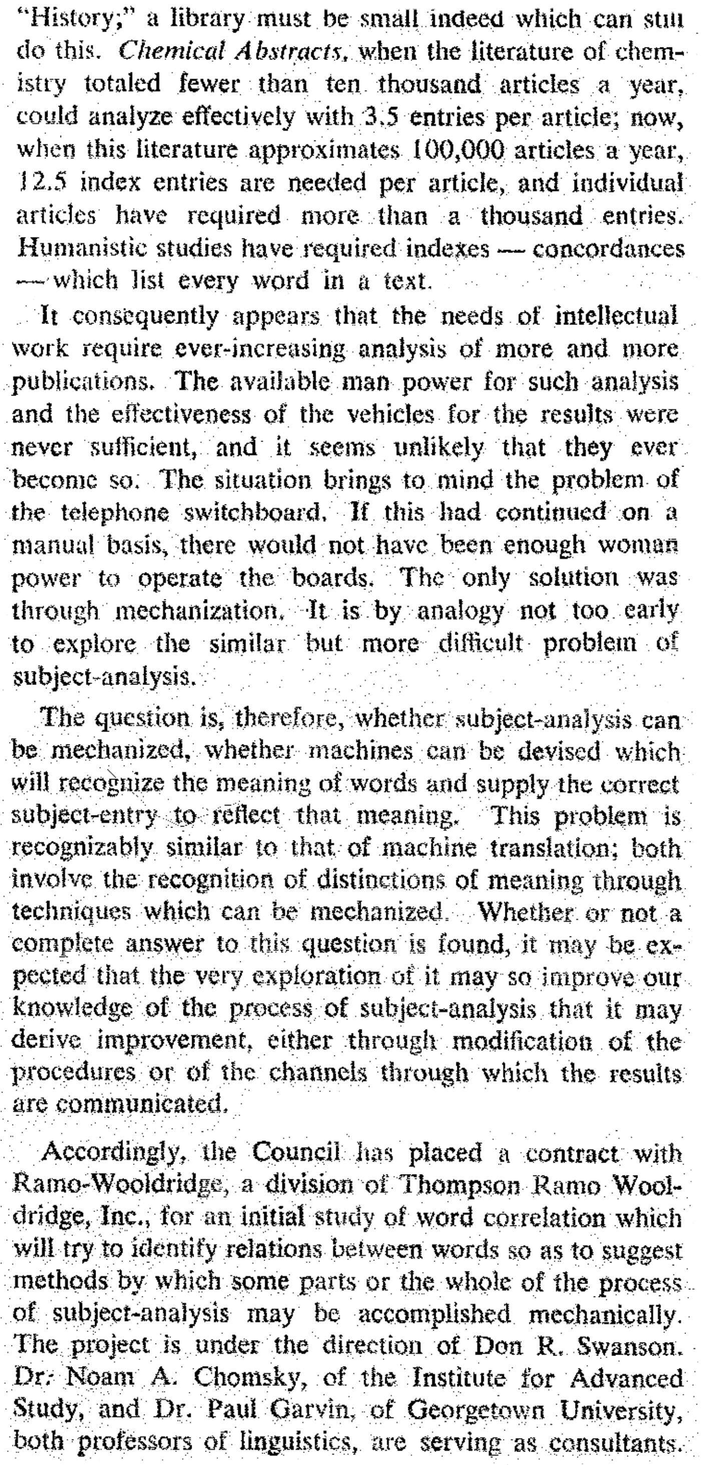


Automatic Indexing

Because the seeker for information never knows, at the commencement of his inquiry, what books or other booklike materials will eventually be found to contain the information he seeks, it is usual to provide clues based on the subject matter of the book. These clues are sometimes expressed in the symbols of a subject classification. Other devices may be employed, too. Typically, however, they are conveyed through words. These words, variously called "subject headings," "index entries," or "descriptors," represent the cataloger's or indexer's judgment both as to the subject-content of the book and as to the verbal route by which an inquirer will seek information on that subject. It is unnecessary to underscore the importance of the subject-cataloger's or indexer's function.

This fact is reflected in the vast amount of subject cataloging and indexing which goes on every day. So exacting and so important is this activity that there are indexing operations carried on even in individual subjects by private professional groups which cost over a million dollars a year.

Meanwhile, the effect of the proliferation of publications has been to require an intensification of the minuteness of the analysis. The Alexandrine Library could adequately arrange the historical books under the simple heading



"History;" a library must be small indeed which can still do this. *Chemical Abstracts*, when the literature of chemistry totaled fewer than ten thousand articles a year, could analyze effectively with 3.5 entries per article; now, when this literature approximates 100,000 articles a year, 12.5 index entries are needed per article, and individual articles have required more than a thousand entries. Humanistic studies have required indexes — concordances — which list every word in a text.

It consequently appears that the needs of intellectual work require ever-increasing analysis of more and more publications. The available man power for such analysis and the effectiveness of the vehicles for the results were never sufficient, and it seems unlikely that they ever become so. The situation brings to mind the problem of the telephone switchboard. If this had continued on a manual basis, there would not have been enough woman power to operate the boards. The only solution was through mechanization. It is by analogy not too early to explore the similar but more difficult problem of subject-analysis.

The question is, therefore, whether subject-analysis can be mechanized, whether machines can be devised which will recognize the meaning of words and supply the correct subject-entry to reflect that meaning. This problem is recognizably similar to that of machine translation; both involve the recognition of distinctions of meaning through techniques which can be mechanized. Whether or not a complete answer to this question is found, it may be expected that the very exploration of it may so improve our knowledge of the process of subject-analysis that it may derive improvement, either through modification of the procedures or of the channels through which the results are communicated.

Accordingly, the Council has placed a contract with Ramo-Wooldridge, a division of Thompson Ramo Wooldridge, Inc., for an initial study of word correlation which will try to identify relations between words so as to suggest methods by which some parts or the whole of the process of subject-analysis may be accomplished mechanically. The project is under the direction of Don R. Swanson. Dr. Noam A. Chomsky, of the Institute for Advanced Study, and Dr. Paul Garvin, of Georgetown University, both professors of linguistics, are serving as consultants.