~~Doctoral Studies in Machine~~ 12 January 1953

Machine Techniques in Scientific Documentation

Introduction Outline of Study (11)

Surge Space

We hear a great deal today about a growing crisis in science. This is not the crisis of the atomic bomb, but rather that of the atomic age wherein greatly increased expenditures on scientific research have resulted, logically enough, in the accumulation of much scientific and technical data. The rate of this accumulation has accelerated so rapidly that we find ourselves in a situation akin to a large city with millions of telephones, but without telephone directories or at least very incomplete ones.

The crisis referred to above is not new. Many foresighted scientists, librarians, etc. predicted in the twenties that we were approaching such a situation. Wartime research and its consequent need for greater information services added to the already accelerated production of research publications. This was further complicated with problems of special handling of confidential information. With this upheaval has developed a new group of professionals who call themselves "documentalists" and more important than this a new concept of "Documentation" -- a term neither new nor presently well defined. Among other things it shall be necessary in this study to review this new concept, trace its history, determine and distinguish between its many definitions and show its relationship to library science, communication and information theory, operations research as well as the general areas of the natural and physical sciences. It would not be overstating the case to imply that "Documentation" may bring a new philosophical approach to science as well as life itself for Documentation deals almost exclusively in the realms of "knowledge" and who knows where this may lead. It is often difficult to distinguish between man's actual knowledge and the ability to communicate that knowledge.

One phase of the documentation problem may be simply stated as indexing. Among other purposes, this dissertation shall attempt to show the importance of indexing in science as well as other disciplines, to define it, and most important,

to indicate the possible consequences of neglecting or ignoring indexing activities. Very briefly indexing activities can be broken into two categories: that which points to what we already know but cannot hope to remember and that which we may desire to know. Suffice it to say that the former problem has already been sufficiently neglected as to cause great concern to historians and scholars in all fields. The latter problem though more neglected, and in the minds of some more important, requires more refined techniques that involve problems of semantics, classification, etc. too involved to discuss in this brief outline. Priority is difficult to establish, but as a minimum "we must be able to find out what we already know." (Chauncey D. Leake, personal communication)

The art and technique of indexing today involves a great complex of activities. There has been and still is a consequent expenditure of human effort that is amazing to behold. With the quantitative increase in man's recorded knowledge this effort has reached huge proportions. New techniques and methods are called for, similar to but different than the problems of mass production, mass education and other quantitative problems. One approach is through mechanization. Can indexing be mechanized? To answer this question it is necessary to analyze present indexing activities and determine what such mechanization would entail and what type of organization will be required. One finds that (important aspects of this problem: (1) machine indexing, (2) machine indexes, and (3) the preparation of indexes by machine.

(1) Machine indexing has reference to the most important, the "intellectual" phase of indexing. This involves the original reading of a document and the decision to indicate the existence of this document under certain headings, names, etc. It is first necessary to ask: can a machine read? We must then also be very specific about the use of the term "read." D'Albe over 30 years ago developed a reader for the blind which translated printed matter into sound patterns. During the war more successful results were obtained by Morton and Flory. We already have such a machine! It is not unreasonable to expect that the same principles with

added refinements will permit printed matter to be read and converted to media such as magnetic tape or punched cards where the information can then be processed. We have yet to learn if the decisions required for indexing are possible by machine. Preliminary studies indicate that suitably designed systems could do an adequate though imperfect job.

On the other hand phases 2 and 3 are possible whether phase 1 be mechanized or not. Assuming that the reading and analysis is still human can we apply machines to the many problems involved in preparing indexes, printed or otherwise. The appended document ~~"Manual of procedures"~~ "THE PREPARATION OF THE CLML BY MACHINE METHODS — A MANUAL OF PROCEDURES." establishes this as a certainty, at least as far as phase 3 is concerned. It is important to indicate here that this problem is by no means of small proportions and in many instances represents the only real obstacle to the preparation and issuance of important bibliographical tools.

Phase 2 is also something not in the future but an activity of several years standing. Previously analysed data can be coded and stored in punched cards, tape or magnetic drum and the consequent "machine index" may be used to great advantage. The work of the Chemical-Biological Coordination Center National Research Council is a good example.

It shall be the purpose of this dissertation to relate and analyze these problems and suggest some solutions.

The studies mentioned above are intended as a prelude to further work on the development of a comprehensive machine indexing system to the world's scientific literature, some of the details of which are included in the appended paper "Current International Scientific Indexes Through Centralized Machine Indexing and its Relation to Standardization of Nomenclature." Thispaper is the result of expanding the scope of the Current List of Medical Literature (see Appendix 1) to the literature of all science.

As indicated in the summary of this paper this plan would hope to (1) help further standardization of international scientific nomenclature; (2) facilitate polyglot indexes to the world's scientific literature; (3) permit indexes

in any one language to the world's scientific literature; (4) help standardize indexes to the individual scientific journals, in the language of the reader; (5) standardize indexes to the individual abstracting and indexing publications; (6) obtain detailed indexing difficult to achieve before; (7) permit simple compilation of indexes to the literature arranged according to any of several classification systems; (8) permit cumulated indexes which incorporate ~~corporate~~ changes in classification systems; (9) provide a check on the coverage of individual articles and journals by the various abstracting and indexing services; and (10) allow for both centralized or decentralized indexing.

Double Space

I. Scientific Documentation

a. Definition and History of Documentation and Review of the Literature

b. The Mechanization of documentation techniques, its definition and history and Review of the Literature

c. Relationship to communication and information theory, operation research, library science, etc.

II. Indexing

a. Definition and History; Review of Literature

b. Importance and Significance for scientific progress

c. Mechanization

1. Machine Indexing

2. Machine Indexes

3. Preparation of Indexes by Machine

III. A Comprehensive Index to the World's Scientific Literature

a. Organization of scientific knowledge

1. Terminology (nomenclature)

a) Standardization

2. Classification vs. alphabetization

b. Centralized machine operations

c. Centralized and decentralized indexing

d. Classified

d. By-products (see items 1 to 10 in summary above.)

(See summary, items 1 to 10.)

Single Space

In closing I should like to quote from F. Donker-Duyvis in his introdution to Bradfordss "Documentation."

"More than half a century ago LaFontaine and Otlet started their lives' work, which at first was connected with the term "Bibliography," but to which gradually the word "Documentation" was attributed. What they developed was of a far wider scope than the establishment of a new technique dealing with printed and non-printed documents.

The great ideal which inspired them, and those who worked side by side with them, was to render accessible the totality of what is crystallized from human thought and to make of it a common treasure of mankind, serving to bring mutual comprehension and to build for peace by the co-operation of all men of goodwill of all nations."

I will feel greatly rewarded to have some contribution towards this goal.