

## The Engineers Joint Council-Battelle Memorial Institute Coordinate Indexing and Abstracting Training Course\*

By JOHN C. COSTELLO, Jr.

Battelle Memorial Institute, Columbus 1, Ohio

and

EUGENE WALL\*\*

Engineers Joint Council, New York 17, New York

Received August 20, 1963

On April 1, 1963, the Engineers Joint Council announced, through a widely distributed press release, the completion of arrangements with Battelle Memorial Institute whereby Battelle would offer publicly on a nationwide basis a course of instruction in abstracting and coordinate indexing. In this announcement, EJC specified that the course developed by Battelle had been granted the Council's exclusive endorsement for a two-year period. The Council further stated that completion of these arrangements with Battelle represented the achievement of the second of three target objectives of the Engineers Joint Council Action Plan.

Before discussing the first and third of these three target objectives in detail, and before discussing how the abstracting and coordinate indexing training course fits in with the other two target objectives, it is pertinent to describe the Engineers Joint Council and identify its objectives. EJC is what might be described as a super-society, in that the only members of the Council are professional engineering societies, such as the American Institute of Chemical Engineers, the American Society of Mechanical Engineers, the American Society of Civil Engineers, and the Institute of Electrical and Electronics Engineers. There are 11 constituent societies, and 18 national associate, regional associate, and affiliate societies, including among others, the American Society for Metals. The Engineers Joint Council was formed for the purpose of performing for the total engineering community and the individual engineering societies those functions which the individual societies could not by themselves do as effectively, as economically, or as well.

One of the efforts which the Engineers Joint Council has felt that it could undertake more effectively, more economically, or better than the individual societies could undertake by themselves is the identification of problems facing the engineering profession in the areas of information storage, retrieval, and dissemination, and the development and implementation of solutions to those problems. The organization within the Engineers Joint Council which is primarily concerned with the identification and solution of problems in the information-handling area is the Information Systems Committee. Members of the Information Systems Committee represent individ-

ual constituent, associate, and affiliate societies. The Chairman of the Committee for the last several years has been Walter M. Carlson, the committee member representing the American Institute of Chemical Engineers. Mr. Carlson's former position with E. I. du Pont de Nemours & Company included responsibilities for improving the utilization of technical information by du Pont engineers and for increasing the utilization of computers for technical and scientific problem solving. Mr. Carlson is now Director of Technical Information in the Office of Research and Engineering in the Department of Defense.

The Information Systems Committee, early in 1962, promulgated what is known as the Engineers Joint Council Action Plan.<sup>1</sup> This plan is based on five fundamental considerations.

The first fundamental of the EJC Action Plan has to do with objective. The objective of the Plan is to improve the efficiency with which practicing engineers, engineering students, and individuals in associated professions can recapture the technical literature pertinent to their current work. A corollary objective is to recover *only* the published information that is pertinent. This fundamental consideration of the Action Plan assumes, first, that the information "flood" will continue, thus making it increasingly less profitable for an individual engineer to read, scan, or even know of the full text of every article or paper of interest to him and, second, that engineers will continue to use words for communication purposes. The second assumption is almost axiomatic, whereas the first is almost inescapable in that 80 to 90% of all engineers and scientists who ever lived *are living today*—studying, designing, developing, building, and above all, *writing*. We deduce that individual engineers, to the extent that they perceive a personal requirement for recorded knowledge, will come to rely more and more upon abstracts and abstracting/announcement services as *initial leads* to information. In such manner they can encompass in a shorter time a much wider range and larger amount of useful or possibly useful recorded information than they could if they relied upon reading or scanning the full texts of information items. This deduction leads to the *abstracting and indexing* phase of the EJC Action Plan, which will be discussed later in more detail.

The second fundamental consideration of the Action Plan concerns language. The use of natural language is

\* Presented before the Division of Chemical Literature, ACS National Meeting, New York, N. Y., September 11, 1963.

\*\* Now Program Manager with Auerbach Corporation, Arlington, Va.

a fundamental method of communication between technical people. The EJC plan is based upon the use of a natural language—English. This fundamental consideration is directed primarily to preventing the interposition of any artificial or synthetic code scheme between the individual who needs information and the information itself.

The third fundamental consideration of the Action Plan has to do with word meanings. Different people use different words to mean the same thing. The EJC Plan is dedicated to working toward procedures which in no way restrict the choice of words by authors but which also ensure that others can find the authors' publications even though they use different words.

From the second assumption stated above, namely that engineers will continue to use words for communication purposes, and from this fundamental consideration regarding word meanings, we deduce a need for developing what is essentially a "translation" medium between the terminologies of different engineering disciplines—in that important new developments often, even perhaps usually, involve the "marriage" of two or more previously unrelated technologies and, accordingly, each engineer is, or should at times be, concerned with technologies other than his own specialty. This deduction leads to the *Thesaurus* phase of the EJC Action Plan, whereby the unification and "translation" technique applicable to diverse engineering terminology (and index entries expressed therein) can be developed.

The fourth fundamental consideration of the Action Plan deals with system compatibility. The procedures adopted in the Engineers Joint Council Plan must be as useful to the individual engineer with a card file in his desk as they are useful to a large, computer-based storage and retrieval system. Basically, the Action Plan when finally and completely implemented must have the capacity for serving the information needs of every individual engineer, not just those who might be fortunate enough to have available to them levels of mechanization for searching more sophisticated than "do-it-yourself" card systems in their desks. Accordingly, the specifications of the Engineers Joint Council on which the training course was developed by Battelle clearly specified that the emphasis should be on fundamentals, such as methods, procedures, and techniques, rather than on devices and equipment for the mechanical storage and searching of indexes.

The fifth and last fundamental consideration of the Action Plan is cost. The cost of abstracting and indexing information at the time of publication is only a tiny fraction of the cost of abstracting and indexing by everyone who stores the information. The cost of creating a comprehensive "Thesaurus of Engineering Terminology" will be relatively small because of the work already done on the "Chemical Engineering Thesaurus" and the steps already taken to ensure compatibility with the new DDC Thesaurus.<sup>2</sup>

We have assumed that individual engineers will come to rely more and more on abstracts and abstracting/announcement services as initial leads to information. If abstracts are to become an increasingly vital tool for engineers in search of recorded knowledge, an economical and quick means of producing and disseminating abstracts

must be found. It is no news that abstracting *can be* an expensive operation. This is particularly true if *informative* abstracts are to be produced, especially by individuals who have to read and understand documents for the *sole purpose* of creating abstracts for them.

Accordingly, EJC has recommended that abstracting be performed by technically qualified individuals who have to read and understand the documents for purposes other than abstracting alone. Further, we have noted that it may be necessary to "settle" for the shorter, simpler, and less costly indicative abstracts, rather than informative abstracts. We have also recognized that "author" abstracts can be useful, although they will often need editing.

The one function which already involves reading and understanding documents is that performed by the editorial staffs of journals. This activity is limited to a relatively few persons who thus can develop among themselves an acceptable degree of uniformity in abstracting. If these individuals add the indicative abstracting and coordinate indexing function to their duties, the additional effort should amount to only five man-minutes or so per article to create indicative abstracts and to provide coordinate indexing terms with roles.

The editing operation, then, is the most economical point for indicative abstracting and coordinate indexing. Further, with relatively little additional processing and with no added cost, the abstracts so created, and the coordinate index terms with roles provided, can be made useful "as is" to other operational organizations "down stream" in the flow system of information dissemination.

Earlier in this paper, it was stated that three target objectives would be discussed. The first of these target objectives has to do with the statement made in connection with the fundamental consideration of cost, namely that the cost of abstracting and indexing information at the time of publication is a tiny fraction of the cost of abstracting and indexing by everyone who stores information. We have stated that the editing operation is the most economical point for indicative abstracting and coordinate indexing. The first target objective of the Action Plan was fulfilled when members of the editorial staffs of the EJC member society publications were trained on the principles of indicative abstracting and coordinate indexing with roles as specified by the Engineers Joint Council. Two such training sessions have been held, one in March, 1962, and the other in March, 1963.

Fulfillment of the second target objective was achieved when the formal training course in indicative abstracting and coordinate indexing was made publicly available starting in April, 1963. The objectives of this training course were, first, to enable information system operating personnel in individual organizations to design and operate their local centers in such a manner that they may make maximum use of the source-created indicative abstracts and term-roles provided by member society publications; and, second, to enable these operating personnel to abstract and coordinate index both their internally generated documents and information obtained from non-EJC sources, utilizing the standardized approaches developed by EJC and taught in the Battelle course. By utilizing the EJC abstracting and indexing

approach on internally generated literature and documents from non-EJC sources, individual organizations will eventually be able to merge with their own work the source-created abstracts and coordinate index terms of approximately 900 technical articles per month from EJC member society publications.

The third target objective concerns the "Thesaurus of Engineering Terminology." Work is nearing completion on the preparation and publication of this Thesaurus. Actual work on its preparation was started on July 1, 1962. The initial steps consisted of collecting, sorting, and screening the individual vocabularies to be refined and edited into the finished product. In all, 18 separate vocabularies were merged, 14 from member societies in addition to the vocabularies of DDC, the "Engineering Index," NASA, and the Engineering Societies Library. In all, about 119,000 terms were submitted; many, of course, were duplicates. The screening and cross-referencing of the assembled lists and the defining of terms selected from them were accomplished by ten subcommittees of an Engineering Terminology Study Committee. Each of these subcommittees was concerned with a specific engineering subject area. The ten subcommittees met for a total of 28 one-week meetings with an average of six technical specialists participating in each subcommittee's work. The committee study phase was completed in July, 1963, and publication of the Thesaurus is scheduled for April, 1964.

The specific purposes of the Thesaurus are: (a) to permit the indexer of documents containing valuable engineering information to index (*i.e.*, describe) more fully, and at different levels of generality and from many technical points of view, the information contained in the documents; and (b) to permit the searcher for information to phrase an inquiry appropriate to the scope and degree of his immediate interests—an inquiry employing all the terms of the retrieval vocabulary which have appropriate meaning and specificity.

In such fashion, the effectiveness of information retrieval can be improved markedly. The indexer can assign to a given document not only index terms brought to mind by the terminology of the author (and of the indexer himself) but also terms which might be used by other technologists—technologists with different interests and viewpoints—to describe or to search for the same document. Similarly, the searcher can phrase his inquiry using terms which *might* have been used to index the document from other points of view.

The thesaurus technique can be flexible and subject to modification to obtain optimum results from an information-retrieval system. For example, if the ratio of inquiries to document accessions is low, the indexers may employ the Thesaurus principally to ensure that they are employing accepted word-forms as index entries (*e.g.*, TESTING rather than TESTS) and to detect and "flag" for consideration as a new vocabulary term any entry which does not correspond to a term already included in the Thesaurus. Consequently, the relatively few inquiries will employ the Thesaurus extensively to compensate for the minimal indexing.

The situation can be reversed to some degree if the output-input ratio is high, although there are strong indications that the Thesaurus will be a useful tool at

output no matter how extensive the indexing effort may be. This may well result from the fact that no set of index entries can ever describe *fully* the information content of a document. Most systems, we suspect, should index documents only to the extent of assigning one indexing term to each useful idea set forth in the document, leaving it to the inquirer to use appropriate alternative terms, near-synonyms, and the like, all of which may be selected by reference to the Thesaurus.

Another form of flexibility is possible when using the Thesaurus to phrase an inquiry. For example, if one wishes only for a "few good references" discussing a particular subject, search terms can be chosen (from the Thesaurus) which will *most precisely* describe the subject of interest, thus screening out most other references which would be pertinent from the point of view of an inquirer wishing complete information on the subject. In other words, the scope of the inquiry can be varied within the limits of the full index vocabulary to suit the needs of each individual inquiry.

We have discussed information retrieval thesauri in detail in an earlier paper.<sup>3</sup>

To serve the purposes described above, the Thesaurus should have several characteristics. It must certainly *list* the terms of the system vocabulary. It must *exhibit relationships* among these terms—relationships of several types, such as synonymy, hierarchy, and relationships which may be synonymous or hierarchical from some points of view but not generally. Finally, it should *define* the vocabulary terms to the extent required and in several fashions. These are the characteristics which have been developed during the work of the Engineering Terminology Study Committee.

The EJC-Battelle course teaches the approaches of indicative abstracting and coordinate indexing using links and roles.<sup>4</sup> The philosophical basis for links in the EJC system remains unchanged from what has been written on this technique in many published papers.<sup>5</sup> The system of roles adopted by EJC and its member societies<sup>6</sup> has evolved through a series of refinements from a system originally developed by the authors in the Engineering Department of du Pont.<sup>7</sup> The du Pont roles were developed from observation and study of similar syntactical control schemes in use at the U. S. Patent Office,<sup>8</sup> the Linde Company,<sup>9</sup> and Western Reserve University.<sup>10</sup> Components of the role system of all three organizations were adapted for use in publishable inverted coordinate indexes for desk use by scientists and engineers. With minor modifications, the original system of du Pont roles was used by the American Institute of Chemical Engineers for providing indexing in term-roles of articles appearing in *Chemical Engineering Progress*.<sup>11, 12</sup> Industrial organizations which use the original system of roles, some with slight modifications, include the Monsanto Hydrocarbons Division Engineering Department, Atlantic Refining Company, Standard Oil of Ohio Refinery Engineering Unit, American Machine and Foundry, Armour Research Foundation, and the Materials Command at Wright-Patterson Air Force Base.

The role system adopted by EJC reflects six years of experience in operating situations. There has been some simplifications and redefinition of roles, primarily to overcome earlier objections that definitions were too

heavily oriented toward usage in chemistry and not adequately oriented toward usage in engineering. The system as now constituted has been tested successfully on information in all fields of engineering and the physical sciences. Organizations which already have adopted and are using the EJC system of roles include Canadian Industries Limited and the B. F. Goodrich Company.

To date, nine of the course sessions have been held in various cities around the country, and between now and the end of the year, an additional 11 are scheduled. Approximately 60% of the course is devoted to practice and drill work in abstracting and in applying links and roles in coordinate indexing, and 40% to background, orientation, and introduction. The fundamental assumption in the course is that coordinate indexing using term-roles is actually and simultaneously indicative abstracting. Accordingly, nine hours are devoted to drill in coordinate indexing, three hours are devoted to abstracting, and eight hours are devoted to indicative abstracting and coordinate indexing simultaneously. Only a rather modest level of familiarity is assumed for the registrants with respect to introductory and background material. Every effort is made to adjust the level of discussion and lecture to fit the over-all background and experience of each group of registrants, and particular attention is devoted to making sure that the less knowledgeable registrants do not get lost early in the course due to insufficient grounding. Course content, methodology, and order of presentation are continuously reviewed in light of comment and evaluations the registrants are asked to submit upon completion of the session.

Finally, Battelle plans to hold a "users conference," later in 1964, for all individuals who have completed the course at any of the sessions and others with "dirty hands" in the actual planning, designing, installation,

and operation of concept coordination systems in user situations. This conference will be for the purposes of exchange of ideas, techniques, and solutions to input and output problems, and generally for discussion of work done in individual situations which might be of benefit to personnel in other operating situations.

## REFERENCES

- (1) "Proceedings of the Engineering Information Symposium," Engineers Joint Council, New York, N. Y., 1962.
- (2) "Thesaurus of ASTIA Descriptors," Defense Documentation Center, Arlington, Va., 1962.
- (3) E. Wall, "Information Retrieval Thesauri," Engineers Joint Council, New York, N. Y., 1962.
- (4) J. C. Costello, "Training Manual and Workbook for Use in Abstracting and Coordinate Indexing Training Course," Battelle Memorial Institute, Columbus, Ohio, 1963.
- (5) J. C. Costello, *Am. Doc.*, **12** (1), 20 (1961).
- (6) "The Engineers Joint Council System of Roles," Battelle Memorial Institute, Columbus, Ohio, 1963.
- (7) J. C. Costello, and E. Wall, in "Studies in Coordinate Indexing" Vol. V, Documentation Inc., Bethesda, Md., 1959.
- (8) D. D. Andrews, "Interrelated Logic Accumulating Scanner (ILAS)," U. S. Patent Office Research and Development Report #6, Department of Commerce, Washington, D. C.
- (9) F. R. Whaley, in "Information Systems in Documentation," J. R. Shera, A. Kent, and J. W. Perry, Ed., Interscience Publishers Inc., New York, N. Y., 1957, pp. 352-383.
- (10) J. W. Perry, A. Kent, and M. M. Berry, "Machine Literature Searching," Interscience Publishers, Inc., New York, N. Y., 1956.
- (11) R. Morse, "Information Retrieval," *Chem. Eng. Progr.*, **57**, 55 (1961).
- (12) B. E. Holm, *ibid.*, **57**, 73 (1961).

---

## Compiling a Technical Thesaurus\*

By T. L. GILLUM

Defense Documentation Center for Scientific and Technical Information,  
Arlington Hall Station, Arlington 12, Virginia

Received May 23, 1963

The development of mechanical methods of storing and retrieving scientific information has brought about an evolution in the form of the traditional indexing authority list. One of the new methods of organizing an indexing vocabulary for efficient use by indexers and searchers has been called the thesaurus concept. A thesaurus is a device for controlling and displaying an indexing vocabulary. The vocabulary is controlled in the sense that it is prescriptive, the individual terms being carefully chosen and appearing as distinct, though interrelated entities. The vocabulary is displayed in such a way that the user

may select all appropriate terms rapidly and accurately.

In the following discussion an attempt will be made to explain how control and display of an indexing vocabulary can most effectively be attained. No effort will be made to compare or contrast systems or vocabularies that use the thesaurus concept with those that use other approaches to indexing. Neither will any attempt be made to defend or condemn the use of the word "thesaurus" in this connotation although there has not been unanimous acceptance of this usage. Furthermore, in this discussion the words "indexing terms" or "terms" will be used to mean words or combinations of words that resemble as nearly as possible the scientific terminology that is normally encountered in the literature. This is done with

\* Presented before the Division of Chemical Literature, 143rd National ACS Meeting, Cincinnati, Ohio, January 14, 1963.