The Abstractor and Indexer*

By ERICK I. HOEGBERG
Research and Development, American Viscose Corporation, Marcus Hook, Penna.
Received January 29, 1962

In discussing the role of literature chemists as abstractors and indexers in industry, I should like first to point out why abstractors and indexers are needed, then discuss what they do, and finally mention some of the kinds of literature they work with and some of the techniques which they use.

I. Why Abstractors and Indexers are Needed.—In industrial scientific research, it is essential that both the long-range investigator and the scientist or engineer concerned with applied research be kept aware of the latest developments in their own and related fields. They also need to be able to learn at will the results of earlier work by others, working within their own companies or anywhere else in the world, which may be applicable to their current research problems. The industrial abstractor and indexer help to fill these literature information needs.

II. What Abstractors and Indexers Do.—The abstractor helps to fill both the industrial research needs which I have just mentioned. The research man can inform himself of new developments in his field by scanning the *titles* of newly published journal articles, government reports and other documents. By so doing he can select the documents which he must read in order to keep up with his field. In many cases, however, the titles of articles do not give sufficient information about the content of an article and the researcher may miss an important paper. In other cases, the titles may be in a foreign language which the researcher does not understand. In either case, if the scientist or engineer can be supplied with abstracts of the documents likely to be pertinent to his work and his interests, he can make more sound decisions as to which documents he should read in their entirety. In many instances, the abstract alone will provide him with all the information he needs and he will not have to read the entire article. If this occurs when the original article was in a foreign language, the need for making an expensive translation may thus be avoided with the aid of a good abstract. For the same reason, abstracts are extremely helpful and time-saving when the scientist is reviewing earlier work to obtain background information or specific information to apply to a current problem.

The industrial abstractor may be called upon to prepare either descriptive or informative abstracts. A descriptive abstract tells in general statements what the document is about, while an informative abstract gives a brief summary of the actual content of the document. Descriptive abstracts are usually brief and are considerably easier to prepare. They are intended primarily to lead the researcher to the original document, rather than to be a partial substitute for it. In preparing an informative abstract, the abstractor attempts to condense the maximum amount of essential information from a document into a minimum of words. In this way he provides a summary or synopsis of the original document.

A competent industrial abstractor must have a broad and sound background in the field of science which is the chief business interest of his firm. In other words, an abstractor working for a chemical company should preferably be a trained chemist or chemical engineer. He must have sufficient experience and maturity to recognize quickly what is new and important in a document, *i.e.*, to separate the wheat from the chaff. Furthermore, he must have a facility for expressing the essential content of the document in clear and concise language. Finally, he should be able to abstract from at least several foreign languages if he works with the published scientific literature of the world. Adding all these requirements together, we find that it takes a highly qualified, top-notch individual to be a good abstractor.

The indexer is chiefly concerned with making information contained in widely scattered documents available for future use. He does this by selecting key words, concepts, terms, or phrases which define individual subjects, or parts of subjects, in the document which he is indexing. The key words, concepts or terms, together with some means of identifying the individual documents to which they refer, are combined with similar entries or subject headings for other documents to form an index. These index words or terms usually are arranged in some logical fashion, such as alphabetically or classified by subject. The index entries and means of document identification may also be so prepared that they can be stored on punched cards, punched tape, or magnetic tape so that the index can be searched rapidly by mechanical sorters or by a computer.

The qualifications for a good indexer are very similar to those for a good abstractor. He, too, must be able to analyze a document for its essential information and then be able to express this information in concise and accurate terms—often from a limited, pre-selected glossary—that best describe the subject content of the document. The indexer, however, often works with the company's internal literature or with the abstracts prepared by the abstractor. Consequently, the foreign language requirements may be somewhat less stringent for the indexer.

III. The Materials and Tools of the Indexer.—The literature with which the industrial abstractor and indexer work may either be the published literature, available to anyone, or the internal technical literature generated within the company for which they work, or both. The published literature may consist of articles published in technical or scientific journals, government research reports, trade pamphlets, patents, etc. Patents, both U. S. and foreign, are an important source of vital information and industrial technological know-how which may never be revealed elsewhere. The internal company literature

* Presented at the Fourth Delaware Valley Regional Meeting, American Chemical Society, January 25, 1962, Philadelphia, Penna.

with which the abstractor and indexer work may consist of laboratory or engineering reports, technical memoranda, market research surveys, patent applications, new product bulletins, *etc.*

The published literature is, in many subject areas, adequately covered by commercial or professional abstracting and indexing services. These services, available to anyone for a price, include *Chemical Abstracts, Physics Abstracts, Electronics Abstracts, Biological Abstracts,* the Uniterm Index to U. S. Chemical Patents, the Derwent Information Service, which provides abstracts of foreign patents in a number of fields of technology, and many other similar services. Some cooperative industry abstracting services, such as that of the American Petroleum Institute, are available only to member companies and are intended to reduce duplication of effort in that particular industry. An industrial firm will not ordinarily do any abstracting and indexing of the literature covered by these commercially available services unless these outside services are too slow or if their output is unsuitable for use by the industrial firm because of inadequate coverage, a different format is required for machine handling, or for other reasons. If the company is working in a new area of science or technology, there may be no commercial services available, in which case the firm is then forced to do its own abstracting and indexing of the published literature. In general, however, industrial firms find it economically advantageous to make use of commercial abstracting and indexing services wherever possible. Consequently, abstractors and indexers in industry are likely to find themselves more concerned with the documents generated within the company than with the published literature.

A company's internal literature presents a somewhat different problem from the published literature. The only abstracts or indexes available are those which are prepared within the company. The internal literature often represents tremendous investments of money and man-hours by the company. If these heavy investments in research and engineering are to be of any present and future value to the company, it is absolutely essential that the firm's internal technical literature be adequately organized by being abstracted and indexed.

The internal literature usually can be segregated into categories having differing degrees of value to the company. Literature items of only transient value obviously do not justify as much time and effort to abstract and index as literature which is of more lasting value. Consequently, several different indexing systems of varying degrees of complexity may be used within a company. Small units of information of temporary value, such as memoranda, will usually not require abstracting and may be filed and indexed by a relatively simple system. Larger items of information, such as research reports, engineering surveys, *etc.*, usually will be abstracted and indexed in considerably greater depth, since they are of more permanent value and represent larger investments of time and money.

Technical memoranda or letters, which generally are brief and deal with only one or a small number of major topics, are often filed in a single location by subject, either arranged alphabetically or by a numerical classification, such as a decimal classification. These small units of information require few, if any, cross-references to other subjects which may be mentioned in the memorandum.

Technical reports and other documents which deal with a wide variety of subjects require more extensive indexing if the valuable information they contain is to be retrievable for future use. Such documents require a more complex system with capabilities for cross-referencing from the viewpoint of any of the individual subjects with which they deal. Here the indexer has a wide choice of systems, and the system chosen within any particular company will be determined by economic factors, by the size and rate of growth of the document collection, by the availability of ready-made systems which can be applied or adapted, and by the training and characteristics of the people who are expected to use the system. For example, a system intended for self-service use by a wide and variable clientele of research personnel must be engineered to be quite simple to use, while a system which is intended to be used exclusively by a small staff of trained documentalists or literature chemists can be made much more sophisticated.

Systems used for company literature in various industrial firms cover a wide range, and in some of the larger firms a number of different systems may be used in different divisions or departments. Many companies rely heavily on alphabetical card indexes. These can be quite simple, using rather general subject headings, or they can be based on more detailed subject headings patterned after those used in *Chemical Abstracts* indexes. When used together with auxiliary card indexes arranged by author, department, project number, and empirical formula of chemical compounds described, this type of index can be particularly well suited for self-service use and can be relatively inexpensive to install and maintain.

Another rather simple type of index is the coordinate index, in which the identification number of a document is recorded under each of the subject terms which apply to that particular document. For example, in indexing a document dealing with the use of magnesium in rocket fuels, the indexer would list the identification number of the document separately under each of the terms: "Fuel," "Rockets" and "Magnesium." The user of this index who is searching for all documents dealing with the use of magnesium in rocket fuels would merely compare the numbers listed under each of the above headings. Those which appeared under all three headings would then represent the desired documents. Coordinate indexes are also well suited for self-service use.

Finally, I should like to touch briefly on machine methods as they affect the industrial abstractor and indexer. The design of indexing and retrieval systems for information is a highly intellectual task. The analysis of documents in order to prepare abstracts and to select the correct index entries is also an intellectual operation, while the steps involved in inserting these items into the information system are largely clerical. Similarly, the formulation of the correct questions to ask the information system in order to obtain the desired information is intellectual, while the actual searching is a largely clerical task. Considerable progress has been made in automating the clerical functions and many machine information retrieval systems are now in routine use. However, automation of the intellectual functions is still in its infancy, although encouraging forward strides have been made in the direction of automatic abstracting and indexing of

documents with the aid of computers. Consequently, it does not appear that skilled human abstractors and indexers are likely to be displaced by machines in the near future.

To summarize, we have shown that abstractors and indexers are needed to make industrial research and development work more effective by helping to avoid unnecessary duplication and by making maximum use of research results obtained elsewhere. We have discussed the work that abstractors and indexers do and lastly we have explored briefly some of the tools and types of literature that they work with.

The Literature Research Chemist*

By STUART M. KABACK
Esso Research and Engineering Co. Linden, New Jersey
Received May 7, 1962

Much has been said and written of the literature explosion which faces today's scientist. It is not at all difficult to see why this situation should exist: it has been estimated that, of all scientists who ever lived, some 80–90% are living today.[1] This army of scientists has nourished a rapidly growing literature, and helped to give prominence to a relatively new type of chemical specialist—the literature research chemist.

Ample evidence documenting this literature explosion is readily available. Price has noted the exponential growth in the number of scientific journals over the past 200 years.[2] The size of *Chemical Abstracts* has grown, over a relatively short period, from 42,000 abstracts in 1948 to 145,000 in 1961. The technical library at Esso Research and Engineering Co. now contains approximately 2,000,000 literature and patent reference cards (representing about 500,000 documents), and this number is growing at a rate of about 100,000 cards per year.

As this flood of information continues unabated, the total of all information, past and current, contained in in the literature mounts to increasingly staggering proportions. The job of the literature research chemist—or, more broadly, the Information Research Chemist—is to stimulate increased and more effective use of this vast store of available data.

The Information Research Chemist is not really a new phenomenon. The laboratory chemist has always doubled as an information research chemist. Today's laboratory chemist may seriously intend to keep up with a number of journals which are likely to contain information related to his interests. However, these journals may easily total several dozens. In practice adequate coverage becomes impossible, as the pressures of achieving and reporting experimental results lead to an encroachment on time set aside for reading. Apparently most chemists feel that they are only being productive when they are in the laboratory. In a recent study conducted by Case Institute, the chemists in several representative groups reported that, on the average, they spend only two hours per week reading the literature.[3]

The adverse consequences of being able to devote so little time to reading are manifold. First, there is a strong inclination to avoid spending time with material not directly connected with immediate interests. Thus, there is little opportunity for the scientist's ideas to be stimulated by work done in more or less closely related fields. Material in foreign languages tends to be ignored. Coverage tends to be superficial. The Case study indicates that frequently the scientist is content to read an abstract or digest, without consulting the original article. Coverage tends to be inaccurate. Obviously, the less time devoted to a job, the greater the possibility of carelessness and error. Finally, it is evident that, as the amount of time allotted to reading diminishes, the chance of overlooking information directly connected with one's interests increases correspondingly.

It must be kept in mind that this discussion so far has referred only to journal literature, and not to patents. In fact, a great deal of significant information first appears in published patents and patent applications—in particular those from such fast-publishing countries as Belgium, Portugal, and the Republic of South Africa. Failure to cover the patent literature often leads to a serious information gap.

The significance of an information gap bears emphasis. It can lead to the unnecessary duplication of effort already expended and documented by others, with a serious waste of time and money. It can result in the chemist's failure to solve his own problems, through ignorance of available ideas, techniques, and results. For the individual chemist, concentration in too narrow an area can lead to serious deficiencies in background when he finds himself working in a new area. Finally, an information gap can cause an organization to miss an important competitive advantage in capitalizing promptly on new trends and ideas.

More and more organizations have recognized the need for technical information groups to cope with the flood of information. A 1959 survey showed close to 100 industrial organizations which had set up technical information groups. These groups range in scope from special libraries which may, in addition, be responsible for literature searching and the publication of abstract bulletins, to integrated multifunctional groups such as the Technical Information Division (TID) at Esso Research.

Three Information Research sections in TID are devoted to Chemicals, Petroleum Products, and Petroleum

* Presented at the Fourth Delaware Valley Regional Meeting, American Chemical Society, January 25, 1962, Philadelphia, Penna.