Technical Information Management in the U.S. Patent Office*

PATRICIA M. McDONNELL
Office of Documentation, U.S. Patent Office, Washington, D.C. 20231

Received July 21, 1969

The task facing the Patent Documentation Divisions within the U.S. Patent Office over the years has not been one of gathering a mass of completely disassociated subject matter into a classification system; it has been rather one of reworking portions of the existing classification system in accordance with an evolving methodology. Now, however, with the development of new and modern methods of information handling and with a growing awareness of national and international responsibilities to the technical and industrial communities, efforts are being directed toward the design and development of nonconventional search systems for shared use under the ICIREPAT (Committee for International Cooperation in Information Retrieval Among Examining Patent Offices) program, and also toward the development and implementation of combined classification-coordinate index search systems.

The Office of Documentation in the U.S. Patent Office is charged with the responsibility of managing a technical information system comprising almost 3.5 million U.S. patents, approximately 7.5 million foreign patents, and an estimated 30 million pieces of nonpatent literature in the patent examiners' search files. In addition, a duplicate classified file of the almost 3.5 million U.S. patents is maintained in the Public Search Room of the Patent Office. It is estimated that patent examiners make approximately 2000 searches per day, and that another 500 searches are conducted each day in the Public Search Room by patent agents and attorneys. These somewhat staggering figures add up to a rather sizable technical information system containing 40 million documents, to which 3000 to 4000 documents are added each week, and in which 2500 searches are conducted each day.

Technical information management in the Patent Office is accomplished chiefly through the work carried out under the Patent Documentation Administrator and the various Patent Documentation Divisions, formerly called Classification Divisions. The professional staff of the Documentation Divisions averages about 70 persons, with a total staff, including technicians and service personnel, of about 135 persons. It has been acknowledged that nowhere in the world is there such a large, qualified, and competent professional staff engaged in creating classifications of technical subject matter. Every classifier has been an examiner for a long enough period of time to be competent in that function, and to appreciate the value of the search system and the problems associated with using and maintaining it. He undergoes a rigorous apprenticeship as a beginning classifier, during which he learns the principles and procedures of developing and administering the U.S. classification system. Both formal classroom and on-the-job training techniques are employed. Recently, this training has been expanded to include a course in the Fundamentals of Automatic Data Processing.

The technicians on the staff are the result of an experiment which was undertaken about four years ago, and which has proved highly successful. A reading comprehension test was administered to a number of nonprofessional employees throughout the Patent Office, and a group of about 15 with the highest scores was selected. This was a group of people who generally either had completed one or more years of college and been forced to drop out for nonacademic reasons, or had actually completed college but had majored in a nonscientific field, and were thus ineligible for professional status as patent examiners. These people were given special training in the principles of the U.S. patent classification system and in some technical areas, and also encouraged to take courses at local universities to augment their previous or in-house training in engineering, science, and languages.

While this formal training was going on, they were assigned to work with classifiers, by whom they were given on-the-job training and experience in performing many of the noncreative tasks involved in carrying out a major reclassification project. The result was that the technicians were able to perform routine, noncreative duties substantially as well as the professional classifiers, at a significantly lower cost. The classifiers were then free to devote a larger percentage of their time to the more challenging, creative aspects of systems development.

The results of the labors of this staff of professionals and nonprofessionals in the documentation area are almost entirely the conventional manual classification search system familiar to many chemical information specialists.

*Presented during the Symposium on Management and Operation of Information Groups and Centers, 4th Middle Atlantic Regional Meeting, ACS, February 13, 1969.

However, within the past few years, some new approaches to search problems have been worked on within the Documentation Divisions, and appear to have promise. More attention will be devoted to these a little later.

The heart of the technical information system in the Patent Office is the U.S. patent classification system—the approximately 350 classes and 67,000 subclasses in the "Manual of Classification," the 15 volumes of "Class Definitions," and the "Index to Patent Classification." This classification system is "an arrangement of all scientific and technical information, to facilitate the selective retrieval of such information when desired. The classification system is intended to make such information available to patent examiners in connection with the examination of patent applications, and to inventors in connection with investigations to determine whether a discovery is patentable or infringes a patent" (*1*).

This definition and statement of intent are from a Patent Office work entitled "Development and Use of Patent Classification Systems," published by the Department of Commerce and available from the Superintendent of Documents. The publication describes in detail 24 different principles which are applied in creating, administering, and maintaining the U.S. patent classification system. Space does not permit discussing them in detail in this paper. However, there is one word in the definition of the U.S. patent classification system which merits particular comment, and that is the word "arrangement." The system is one for placing documents in filing cabinets (called shoecases) in such a manner that the information can be located in the shortest time with the minimum effort. It is basically a filing system, as opposed to an indexing system. Many of the 24 principles are concerned with techniques for minimizing the number of copies of each document which must be placed into the system. Much of the criticism of the system comes from people who are attempting to use it as an indexing system, for which is required assignment of greater numbers of subclasses per document than would be appropriate for Patent Office purposes. In the U.S. patent classification system, a patent occurs in an average of 2.5 subclasses, which is obviously not enough access points if the classification is to be effective as either a subject index or a coordinate index.

In recent years, some other techniques and newer approaches to coping with information management problems have been employed in the Patent Office. Of particular significance is the ICIREPAT program.

ICIREPAT is an acronym for Committee for International Cooperation in Information Retrieval Among Examining Patent Offices. It was conceived in 1961 when the U.S. Patent Office was celebrating the 125th anniversary of the Patent Act of 1836. Representatives of the patent offices of nine countries and the International Patent Institute of The Hague met in Washington to discuss mutual search problems in the various offices. It soon occurred to the group that some sort of cooperative effort to help solve search problems and to avoid duplicate effort would be of substantial benefit. An organizational meeting was held in Munich in 1962 and ICIREPAT was born. In addition to the original conferees, representatives joined from the patent offices of five additional countries, and also Euratom. In the last six years, representatives have been added from a number of additional countries and treaty organizations, bringing the total membership to about 25 patent offices and international treaty organizations.

In 1964, after much hard work, the first program was adopted by the members at the 4th annual meeting in Washington. This program consisted of two phases. The first involved the adoption for international shared use of a group of existing systems which had been independently developed in the various countries and were being used as search tools in those countries. The selection of existing systems was for the purpose of getting the program off to as quick a start as possible.

The second phase of the program involved the establishment of procedures for cooperative, international, shared development of new systems. An Advisory Board for Cooperative Systems (ABCS) was established, whose task was the coordination and supervision of the program.

The ABCS defined specific stages through which a proposed system must sequentially progress along the path to international operation. First, for new systems proposed for shared development, the following five stages were set forth:

1. Notification stage
2. Term list stage
3. Indexing trial stage
4. Search trial stage
5. Operational indexing stage

Notification Stage. In this stage, any member office may signify to the ABCS its intention to act as originating office for a particular field of technology. This notification will carry with it a reasonably precise identification of the subject field, identified in natural language, as well as by reference to the national classification and the international patent classification; a statement as to the approximate size and nature of the document collection; the approximate search activity, stated as the number of applications filed per year; and the expected completion date of the first draft term list.

The ABCS chairman then designates a monitor for the field, and notifies all members, requesting from each a statement as to that office's interest in participation and information as to the size of the file and search activity in that country.

Term List Stage. The draft term list is circulated to all member offices which have indicated an interest in participation. At this stage, the members have the opportunity to criticize and comment on the sufficiency, definition, scope, and detail of the term list. This gives an office an opportunity to orient the term list toward any peculiarities in its own practice. The various comments form the basis for a revision by the originating office.

Indexing Trial Stage. Here the term list is further tested to be certain that all ambiguities and deficiencies have been eliminated and to establish an indexing consistency measure for the field.

Search Trial Stage. A sample file is constructed—for example, by making a random selection of about 500

documents—and is then used for conducting a number of searches based upon current applications or recently issued patents.

Operational Indexing Stage. At this point, the developed field has been approved by ABCS for shared use by the members. All participating offices then share in the indexing work.

Recently, a sixth stage, operational searching and updating, has been established. Five systems have entered this stage, including three in the chemical arts—alloys, lubricants, and layered products—although none are fully operational in the U.S. office as yet.

When an originating office would like to offer an existing system for shared use, the procedure is quite different. Since members cannot recommend substantial changes in existing systems, the originating office is required to submit sufficient data regarding the system and its characteristics to a designated monitor, who studies it and makes a recommendation with regard to its acceptability under existing criteria. The Stage 1 notification is also accompanied by the same data. If such a system is accepted, it in effect then enters into the equivalent of Stage 5 of the procedure for new systems.

Under its new organizational rules adopted in September of last year, ICIREPAT will be headed by a Technical Coordination Committee (TCC) composed of representatives from eight participating countries, six of which are those whose national offices receive, according to the latest available statistics, the greatest number of patent applications per year. This committee will organize, supervise, and coordinate the activities of special working groups called technical committees. At the first meeting of the TCC in April 1969, the following six technical committees were established: Retrieval Systems Design and Testing, Technical Fields: Forward Planning, Advanced Computer Techniques, Microform, Patent Format and Printing, and Systems Implementation.

What is the current status of the ICIREPAT program? About 57 new and existing systems have been proposed by the various member countries to date. About 49 of these systems are in some stage of development; the remainder are inactive, either for lack of sufficient interest by the various members or for failure to obtain approval by ABCS. The approximately 49 live systems comprise a total of close to 1 million patents. This is not to say that the United States will necessarily adopt as a primary search tool all of the proposed systems, but it is at least interested enough to consider them for adoption if they appear to be beneficial when advanced enough to be fully evaluated.

Several of the systems are at present operational in the U.S. office, including the steroid file and the organometallic file in the chemical arts. These systems originated in this office. Both are presently being revised in accordance with requests made by several other member countries, and reindexing of the steroid file is now beginning. The U.S. office has originated nine systems and expressed interest in an additional 25 systems.

Certainly, the trend of the times is toward international cooperation, and especially so where all parties concerned can benefit from mutual labors. International cooperation among patent offices is being evidenced not only through the ICIREPAT program, but also through efforts directed toward the adoption of the proposed Patent Cooperation Treaty. The U.S. Patent Office has high hopes for these cooperative efforts, which can help to reduce the staggering and ever-increasing search loads imposed on its examiners.

The above is a very sketchy glimpse of a complex program. It is really too soon to anticipate all of the implications of this program. Certain extremely important factors are still not completely clear. It is not known, for example, just what the manpower requirements are to implement the systems. Also, truly reliable procedures have not been developed for evaluating the worth of suggested systems before commitments are made to index thousands of documents. The U.S. office is optimistic that these and other problems will be resolved, and much benefit can be derived from this program.

The next significant, relatively recent approach to coping with technical information problems in the Patent Office is what is called the "dual system," which is actually a conventional manual classification search system combined with a supplementary mechanized coordinate index search tool (*2*).

Several years ago, in what was then the Office of Patent Classification, work was begun on the reclassification of Class 167, Medicines, Poisons, and Cosmetics (*3, 4*). The conventional approach to developing a manual system of hierarchical classification for patent searching was not satisfactory in so far as this particular class was concerned. The problem stemmed from the varied nature of the disclosures and searches within this class.

Conventional hierarchical classification presents serious difficulties to the classifier when plural bases for search are employed by the examiner. Searches during prosecution of patent applications in Class 167 fell into three distinct categories: chemical compositions (mixtures) disclosed as having a pharmacological, biological, or cosmetic utility; methods of preparing such compositions; and methods of utilizing chemical compositions or compounds for pharmacological, biological, or cosmetic utility. Since methods of preparing compositions are generally classified with the compositions themselves, the two obvious bases for classification of patents in this class were chemical structure and utility. Both are significant factors in searching; however, selection of one as the primary basis of classification would virtually prohibit searching based on the other in a purely conventional classification search system.

Reclassification of the patents in Class 167 has been effected by preparation of a conventional manual classification search system supplemented by a coordinate index system. Upon consideration of the facts that the aspect of chemical structure is common to all three types of searches, while utility is significant in only a major portion of them, and also that indexing of chemical compositions for mechanized search systems is extremely complex, it was decided that the primary basis for development of the manual search system would be the chemical structure of the disclosed and claimed compositions; the supplementary coordinate index system would provide for searches relating to the pharmacological and other biological functions or utilities. Cosmetic utility is excluded from the coordinate index search system, but is included as a special area with a high priority in the conventional classification schedule.

Concurrently with the development of the conventional hierarchical schedule, six descriptor lists were prepared for the various utility-connected facets of disclosed information which were deemed to be of interest to the examiner. A thesaurus was also compiled to assist the indexer or searcher in selecting appropriate descriptors. The conventional subclass schedule functions as a seventh descriptor list for chemical information.

In the process of creating the manual portion of this class, assignment of the patent documents to appropriate subclasses as originals and cross references was accomplished in the usual manner by placement of working-number subclass designations on the patents themselves. After consideration of each document for placement in the chemical classification schedule, the classifier completed a coding sheet for the utility data disclosed. As many terms were selected and noted on the coding sheets as necessary to index fully the pharmacological and biological data disclosed in the patents. It is estimated that the step of indexing a patent for the coordinate index system added an average of only five minutes to the time normally spent for the classification of a patent on the final pass. Thus, for a relatively little expenditure of additional time and money, the retrieval capabilities of the conventional system were greatly expanded. Because of using the conventional classification subclasses as chemical descriptors in the coordinate index segment, it was not necessary to devise a complex machine code for the chemical structures, and the time for encoding the chemical information was substantially reduced.

The fact that the system is not complex and that it is relatively easy to encode the documents means that large segments of the arts may be encoded and searched. The new Class 424, which replaces the old Class 167, contains in the mechanized segment approximately 26,000 U.S. and 25,000 foreign patents, for a total many times larger than any existing mechanical system that we have actually in operation for searching patents.

Another innovation has been introduced into the combined system, this one in the handling of cross reference patents which are originals in Class 260, Chemistry, Carbon Compounds. In the old Class 167 system, there were about 14,000 patents originating in Class 260 which had claims to new compounds and methods for their preparation and were cross-referenced to Class 167 because they contained disclosures of compositions and/or utility which were subject matter for Class 167. Approximately 9000 of these Class 260 patents did not contain any disclosure of compositions for Class 167, but only of utility. In the new combined system, cross-reference copies of these patents, exclusive of those in the U.S. steroid mechanized search system, were not placed in the new Class 424, but were crossed into Class 260, Subclass 999, and indexed in the coordinate index system only. From standard Patent Office classification files, information as to the subclasses in Class 260 in which these patents occur, both as original and as cross reference, was obtained and entered into the coordinate index system. These subclasses are then used in formulating search queries in much the same way that Class 424 subclasses are used in searching patents occurring in that class as originals and as cross references. Because cross reference copies of these 9000 patents did not need to be placed in the new Class 424 conventional search file, these patents were indexed in approximately one-half the time required for other patents which also had to be reclassified. Patents claiming steroid compounds only, without any disclosure of composition, are searched by using the steroid mechanized search system.

The combined classification-coordinate index system is partially operational at this time and is expected to be fully operational in the Examining Corps and in the Public Search Room by mid-1970. At that time, the searcher will have considerable flexibility in making various types of searches. A search for a new composition or method of preparing a composition will be made through the conventional manual classification system alone, and will be somewhat more efficient because of the exclusion from the manual search file of those Class 260 patents which disclose Class 424 subject matter relating only to utility and not to composition. Thus, while the former Class 167 search system would now contain approximately 29,000 discrete U.S. patents, the new manual search file contains only 20,000. Since many of these 9000 documents which are no longer in the manual file would have been crossed into two or more subclasses, the new Class 424 file size is actually reduced by approximately 15,000 documents.

The search for utility will be through the coordinate index segment, and will be variable by selecting any combination of facets desired and by exhibiting a genus-species relationship within those categories which are hierarchical. In addition, the search may be limited to a given combination of structure and utility by specifying the conventional subclass number, either Class 424 or Class 260, of a desired structural characteristic as one of the categories in the question.

Initially, implementation of the coordinate index search system will be with Termatrex optical coincidence equipment. However, all of the data are stored in standard 80-column punched cards, and it is anticipated that searching may be done by computer when the document file becomes so large as to make optical coincidence searching unwieldy. Document retrieval will be via a file of aperture cards stored in Termatrex accession number sequence.

This system helps bridge the gap between complete mechanization and manual systems. At such time as machine searching of chemical structures, including Markush structures, and compositions or mixtures becomes feasible in the Patent Office context, the manual portion of the system could be replaced by a mechanical portion for composition searching. The implementation with Termatrex has a further subtle advantage of gradually introducing the examiners to machine searching.

Two other major reclassification efforts are at present being reclassified in the Chemical Documentation Division, Class 99, Foods and Beverages, and Class 23, Inorganic Chemistry. In both of these projects, the possibility of applying the techniques used in the Class 167 project is being investigated. It is planned that dual systems will be created for these projects as well. The technique has been employed on an experimental basis in some very small, active areas of larger classes, and the preliminary results are very encouraging.

In conclusion, a few comments about the problem of document retrieval are in order. The U.S. patent

classification system is a purely manual search system in which the functions of search and retrieval are combined into one operation. A patent examiner scans the documents in a given subclass or group of subclasses and pulls out those four or five documents which he wishes to take back to his desk for further study and for use in writing his action on a pending application before him. The newer techniques described above for managing technical information in the Patent Office are being employed for performing the search function only, or for identifying those four or five documents, frequently embedded within 40 or 50 drops from a single search, which the examiner wants to take back to his desk. However, until such time as the difficulties in retrieving the desired documents and getting them into the examiner's hands effectively and efficiently can be overcome, it is felt that large-scale replacement of the present U.S. patent classification system will not come about.

LITERATURE CITED

(1) U.S. Department of Commerce, "Development and Use of Patent Classification Systems," p. 1, 1966.
(2) McDonnell, P. M., "A New Approach to Mechanized Information Retrieval in the U.S. Patent Office," *J. Pat. Office Soc.* **50**, 651-8 (1968).
(3) Vincent, H. S., "A Combined Manual and Mechanical Search System," in *Information Retrieval Among Examining Patent Offices*, Proceedings of the Fifth Annual Meeting of ICIREPAT, 1965, H. Pfeffer, Ed., pp. 535-42, Thompson Books, Washington, D. C., 1967.
(4) McDonnell, P. M., and H. S. Vincent, "A Combined Manual-Mechanical Search System for Patents Relating to Medicines, Poisons, and Cosmetics—Present Status," in *Information Retrieval Among Examining Patent Offices*, Proceedings of the Sixth Annual Meeting of ICIREPAT, 1966, H. Pfeffer, Ed., pp. 275-8, Patent Office, U.S. Department of Commerce, Washington, D. C. (no publication date given).

Transfer of Security-Classified Information*

A. GREGORY ABDIAN and PAUL KLINEFELTER
Defense Documentation Center, Alexandria, Va. 22314

Received May 26, 1969

The safeguarding of security-classified information and the systematized transfer of information are inseparable but fundamentally conflicting requirements of the Defense Documentation Center (DDC) missions and activities. DDC experiences complex and costly burdens to achieve and maintain a realistic balance between information control and information dissemination. These impacts are related to interfaces with sources and users, to in-house processing and storage of documents and data, to announcement media, to data retrieval and miscellaneous products and services, and to the development of advanced systems and capabilities for improved information transfer. Areas of special difficulty or new interest are described and estimates of the cost impacts of processing security-classified information are summarized with respect to various processing functions.

As the major centralized Department of Defense facility for the transfer of scientific and technical information, the Defense Documentation Center (DDC) deals with large amounts of classified information. DDC is a major field activity of the Defense Supply Agency, and supports research, development, test, and evaluation programs being conducted by or for the Department of Defense. DDC collects, processes, stores, announces, retrieves, and supplies information in all the scientific and engineering fields of interest to the Defense Department and its contractor family. Its Work Unit Information System encompasses the storage and retrieval of data about on-going research and development. DDC's Technical Report System deals with documented records of completed work. The agency's information and documentation systems use a large computer facility which, on the one hand, permits great service speed and manipulation of huge amounts of information but, on the other, compounds the difficulties of safeguarding classified information by its very efficiency in providing such material.

The thousands of Defense-sponsored RDT&E facilities, laboratories, and test stations, both in-house and extramural, are simultaneously sources of input to DDC and users of DDC outputs. Security controls imposed at the point of information generation or upon transfer to DDC also apply when data and reports are provided to users by DDC.

DDC is in effect an information wholesaler, and the classified portion of this information moves in a restricted path within limits prescribed by a common set of regulations. DDC itself rarely classifies information, and then only internally-generated bibliographies and reference tools. However, as an element of the federal government, DDC must observe established policies for the safe-

* Presented before the Division of Chemical Documentation, 4th Middle Atlantic Regional Meeting, ACS, Washington, D. C., February 13, 1969.