Searching for Chemical Information in the Patent and Trademark Office†

PATRICIA M. MCDONNELL

Patent and Trademark Office, Washington, D.C. 20231

Received April 20, 1977

Searching in the Patent and Trademark Office is characterized by an extraordinarily large search file, rapid file growth rate, and high search activity. A manual access search system, the U.S. Patent Classification System, has proven over the years to be an effective, economical tool for meeting the particular requirements of examiners' patentability searches and searches of the general public. Advantages and weaknesses of this system, and efforts to overcome the latter, are discussed. Other chemical search tools, such as *Chemical Abstracts* and on-line bibliographic data bases, presently being used by the Patent and Trademark Office are also mentioned. Present and planned future efforts center around the use of a computer-controlled microform search system (CCMSS) which is described in detail.

Searching for chemical information for the determination of patentability is based on requirements flowing from laws passed by Congress and from decisions handed down by the Courts. The present statutory requirements for patentability are set forth in Title 35 of the United States Code. In order to be granted a patent, an applicant must make a complete disclosure of a new and useful process, machine, manufacture or composition of matter, or any new and useful improvement thereof; further, the differences between the subject matter sought to be patented and the prior art must be such that the subject matter as a whole would not have been obvious at the time the invention was made.[1]

Searching in the Patent and Trademark Office (PTO) today consists, for the most part, of manually shuffling through piles of paper arranged according to the U.S. Patent Classification system. This system provides for arranging all publications containing scientific and technical information in such a way as to facilitate selective retrieval on demand. For this purpose, 355 broad classes of technological information have been established, and these have been further subdivided into approximately 78 400 official subclasses and 17 000 unofficial subclasses. The annual growth rate is about 2500 official subclasses.

Although the U.S. Patent Classification System serves a variety of functions, it is designed primarily to provide for patentability searches by patent examiners; hence, the basis for division or placement of documents in the system must be chosen in light of the appropriate statutes and relevant court decisions pertaining to patentability. The principal basis for classification in the PTO is utility, in the sense of proximate function, effect, or product. A thorough discussion of the system and its underlying philosophy is beyond the scope of this paper; however, such information may be found elsewhere in the literature.[2-5]

The approximate size of the patent examiners' search file today is shown in Table I. Patents are classified and copies placed in one or more subclasses according to the inventions claimed and, optionally, other subject matter disclosed in them. Office-wide, the average number of classifications (copies) for each U.S. patent is close to three, i.e., one original classification and two cross-reference classifications. The average for chemical patents, however, is probably closer to four or five.

There are no figures available as to the percentage of these documents which are in the chemical fields. Estimates vary from 25 to 40% of the patent documents, depending upon the definition that one gives to the term *chemical*. A higher percentage of the nonpatent documents are known to be in the chemical classes.

Table I. Size of Patent Examiners' Search File

4.2 million original U.S. patents	
6.2 million official cross references	
1.6 million unofficial cross references	
9.5 million foreign patents	
1.0 million nonpatent items	
Total	22.5 million documents

Table II. Annual Updating of the U.S. Patent Classification System

Reclassification of Existing Patents	New Patents Added
200 000 original U.S. patents	75 000 original U.S. patents
400 000 cross references	175 000 cross references
320 000 foreign	280 000 foreign
920 000 total	530 000 total

In addition to the file for examiners, the PTO maintains a separate file in its Public Search Room in Arlington, Virginia, for use by patent attorneys and agents and any other persons who may wish to search the patent literature. This file is not as complete as that for examiners, however, in that it contains only original U.S. patents and official cross references. The Public Search Room is open week days and staffed with search assistants to help persons unfamiliar with the U.S. Patent Classification System to identify and locate subclasses pertinent to their search needs. The average number of persons who use this facility each day is 300.

The search files are growing at a rapid rate. In addition, the Classification system is constantly undergoing modification and revision to meet the requirements of the new technology being claimed in applications for patents. Table II shows the annual updating estimates for patent documents. A significant proportion of the reclassification effort represents the incorporation of what were formerly unofficial subclasses into the official system.

To complement these files arranged according to the U.S. Classification System, the Patent & Trademark Office maintains a comprehensive Scientific Library for use by examiners and members of the public. The holdings include 250 000 monographs (books), 90 000 bound volumes of periodicals, and 12 million foreign patents arranged numerically by country, either as bound volumes or on 16-mm roll microfilm. The Scientific Library subscribes to approximately 2000 journal titles annually. Use of the Scientific Library materials is highest among chemical examiners, in recognition of which the Library maintains a separate facility in the midst of the chemical examining operation for maximum convenience.

† Presented in the symposium on "Meeting the Challenges of the Changing Patent Literature", Division of Chemical Information, 173rd National Meeting of the American Chemical Society, New Orleans, La., March 21, 1977, and 11th Middle Atlantic Regional Meeting of the American Chemical Society, Newark, Del., April 21, 1977.

Search activity in these voluminous files is very high. The PTO processes over 2000 applications each week. Each application typically receives a thorough search for the first action (letter to the applicant from the examiner on the merits of the application) and a second or updating search on each subsequent (usually one) action. Again, there is no breakdown on the number which are chemical, but one may assume about 40–50% require at least some searching for chemical information. Thus, about 1800 searches (i.e., 900 first-action searches and 900 updating) are made each week by examiners. Another 1000 or so are made in the Public Search Room.

The searching processing is basically one of rejection. Typically an examiner goes through one or (usually) more subclasses containing several hundred documents each, looking at the front page of each document. Based upon what appears there, the examiner rejects most of them as not being useful in examining the application under consideration. If any of the documents appear particularly promising, based on the front page or on knowledge of the contents from past experience, the examiner may read other parts of the document. Those of particular interest (usually 10–15) are removed from the file and taken back to the examiner's desk for further study and use in preparing the action.

The U.S. Patent Classification System has proven over the years to be an effective, economical search tool in an organization with an extraordinarily high volume of searching and large file to be searched. A number of key features contribute to its viability. First, searching entails not just looking for documents disclosing what is being claimed, but also looking for documents which disclose something close to what is being claimed—documents which, taken singly or in combination, would make it obvious to a reader with ordinary knowledge of the art to do or make what is being claimed. Browsing through large numbers of documents in a manual system lends itself to meeting the requirements of this type of searching, as well as to keeping searchers apprised of the current state of the art.

A second reason why the manual access system is particularly appropriate in the examination process is that it combines the dual functions of document identification and document retrieval. On browsing through a subclass, the examiner not only identifies those documents potentially of use in replying to the applicant but also physically removes them from the file. When the examiner finishes searching, not only have the needed documents been identified, but they have also been retrieved.

A third advantage of this type of search tool is that it is amenable to having portions of it revised or rearranged in response to sudden shifts in technology. These portions may be as small as a single subclass which has suddenly grown to many hundred of documents or as large as several entire classes containing hundreds of subclasses which no longer meet the search needs as dictated by new technological developments and by legal decisions rendered by the courts.

Along with these favorable attributes of the manual search system, there are, of course, certain inherent weaknesses. A single facet (e.g., chemical structure or biological utility) must be selected as the basis for construction of a hierarchical classification schedule, and searches relating to other facets are very difficult and in some cases impossible from a practical standpoint. Second, the paper file of 22.5 million documents takes up a considerable amount of space, and space is becoming a critical resource in the PTO. Third, the paper file is subject to lack of integrity as documents are being removed and taken to examiners' rooms for study. Finally, substantial clerical processing is required to refile documents which have been removed, to add new documents as they are acquired, and to build a new paper file following the professional effort by classifiers in revising the system.

Efforts at automation of searching in the PTO began over 20 years ago but have met with little success. The earliest attempts centered primarily around development of coordinate index systems for selected groups of related subclasses, primarily in the chemical arts, in which the stated aim was to extract and code all of the information in the documents, so that the automated system could be used as a replacement for the manual access system. These systems met with little acceptance by management because of their high input costs and by examiners because the system provided them only with lists of numbers rather than copies of documents. There were several exceptions, however, in arts where the manual file was in need of reclassification and more time overall had to be spent by the examiner with the manual search than with the automated system and its onerous requirement for locating paper copies. These exceptions included systems in the fields of Steroids (still in use today as a primary search tool in the U.S. and other patent offices and by some searchers in the private sector), Organometallics (still used as a supplemental search tool), and Organophosphorus Compounds (formerly used as a primary search tool for one very large subclass but discontinued after it was subdivided into over 70 smaller ones).

In an effort to reduce the development costs of deep indexing costs, the PTO initiated a cooperative effort called ICIREPAT (International Cooperation in Information Retrieval Among Examining PATent Offices). A major part of this effort centered around the Shared Systems Program, in which emphasis was placed on the cooperative development of coordinate index systems and exchange of indexing information for use in searching within each office as its own national needs dictated. The Shared Systems Program has not been as successful as some offices had hoped, but it is still in operation at a reduced level. The procedures for development of systems and some of the underlying reasons for limited acceptability have been described elsewhere.[6,7]

A second approach to cutting development costs which was undertaken in the PTO was the "dual system" concept. This approach was first employed in the development of the present Class 424 (Drug, Bio-Affecting and Body Treating Compositions) and involved the concurrent development of a conventional manual classification search system based on chemical structure and a supplementary coordinate index search tool for pharmacological and biological activity.[8] As with the earlier deep-indexing systems, however, this system was used very little because of the lack of a suitable means for implementation.

In the late 1960s and early 1970s, in view of this lack, system development efforts (other than those under the aegis of ICIREPAT) were limited to the development of small, coordinate index systems to supplement the manual search system in very limited areas of technology, usually involving only about 2000–3000 U.S. patents. The most successful of these were in nonchemical arts in which chemical information was significant. For example, in applications relating to welding processes, a significant number of claims relate to the chemical composition of the filler material used in the welding process. The classification schedule for this art is based on process steps and so does not provide a convenient tool for searches relating to filler composition. An indexing scheme for coding the chemical constituents of the compositions was developed, the patents were indexed, and a deck of Termatrex optical coincidence cards was given to the examiner, along with a retrieval file of paper copies arranged in accession number order. With the aid of this tool, the examiner is able to search for specific filler composition in a few minutes.

Concurrent with this development of small systems in limited areas, a sizable effort was launched in 1969 in the field of

computer-aided classification (CAC) with a view toward making maximum use of the full-text database being produced as a by-product of the conversion of patent printing from hot lead to photocomposition. This effort eventually became known as Project POTOMAC (Patent Office Techniques Of Mechanized Access and Classification) and was expanded to include not only classification but also search and retrieval. In 1972 work was suspended and an evaluation team from the National Bureau of Standards concluded that the goals of Project POTOMAC were not achievable within the present state of the art.

In addition to its in-house efforts, the Patent and Trademark Office has over the years looked for additional search tools available from other sources which the examiners might find useful in their work. *Chemical Abstracts* is searched regularly in all applications involving new chemical compounds, compositions, and processes, and frequently for other chemical cases. To facilitate searching foreign language patents, English language abstracts obtained from Derwent are attached to the front page of such documents before they are placed in the search file. INSPEC, through its Patent-Associated Literature service, selects, abstracts, and classifies nonpatent items for placement in the search files. During the last year examiners have also begun searching the various files available through System Development Corporation and Lockheed. The utility of these searches is presently undergoing evaluation. In many other cases, however, commercially available systems have not served to provide the examiner with sufficiently better art than he finds in his own search file to justify the added time and expense involved in searching these systems and attempting to locate copies of the documents thus identified. The non-use of such systems should not be construed as a rejection of such systems because they do not come up to the quality standards the Patent and Trademark Office tries to maintain in searching but only as a rejection because their use is not cost effective, in that they cannot be used to replace anything the examiner is already doing.

In order to attempt to overcome some of the problems of space and file integrity for the classification system and access to documents for coordinate index systems, the Patent and Trademark Office proposed in 1973 to undertake two experiments in the field of computer-controlled microform search systems (CCMSS): one to be directed toward simulation of what the examiner now does in searching his manual access classified file, and the other to combine the searching of a coordinate index database and the display of documents identified in the search.

Because of the limited funds available at that time, only one of these experiments was undertaken. In June 1974, a contract was let to Image Systems, Inc.,[9] for the installation of a computer-controlled microform search system (CCMSS) for searching coordinate index files. This system was installed in April 1975 and has been enthusiastically received by both patent examiners and searchers from the private sector.

Encouraged by the initial experience with use of the CCMSS for searching indexed files, a new contract with Image Systems was let in June 1976 to upgrade the present software and hardware and to undertake the other of the two experiments by addition of two large files (one of them in the chemical arts) for searching by the U.S. Classification System, i.e., to enable the user to do sitting at a microform terminal what he now does standing at a paper filing case.

The CCMSS consists of a number of Questicon® units (four at this time), all of which are under control of a single minicomputer. Each Questicon unit, shown in Figure 1, consists of a microfiche display device and a keyboard through which the user communicates with the minicomputer and display device. The microfiche (Figure 2) are clipped with

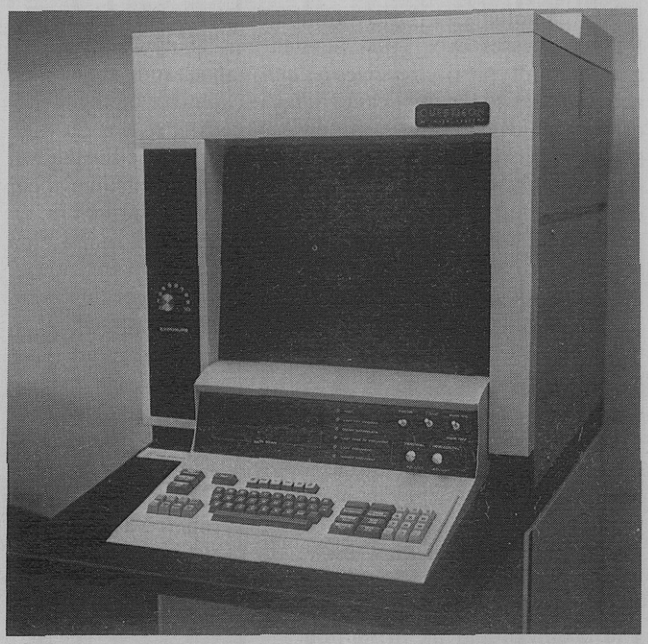

Figure 1. Questicon® micrographic storage and retrieval unit.

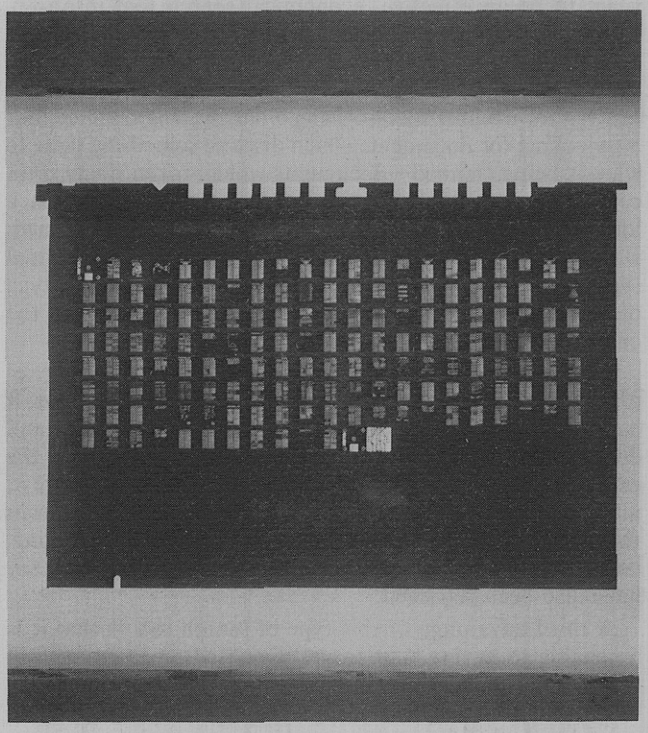

Figure 2. Clipped microfiche.

a binary coded metal strip. Images on the fiche are at 48× reduction, and each fiche can store up to 273 images. The fiche are stored in a carousel (Figure 3) in the base of the Questicon unit. The carousel can store about 750 fiche, so that the capacity of a Questicon unit is about 200 000 pages, or 28 000 U.S. patents. Average access time between patents is 2–3 seconds. Each Questicon unit, in addition to containing pages of documents, also contains pages of instruction and informational material necessary for using the system. Adjacent to the minicomputer there is a teletypewriter by means of which examiners are able to obtain printouts of their search results.

Step-by-step instructions are presented to the searcher by means of the series of INSTRUCTION messages which are displayed on the microfilm screen. Associated with each INSTRUCTION message is a HELP message which contains

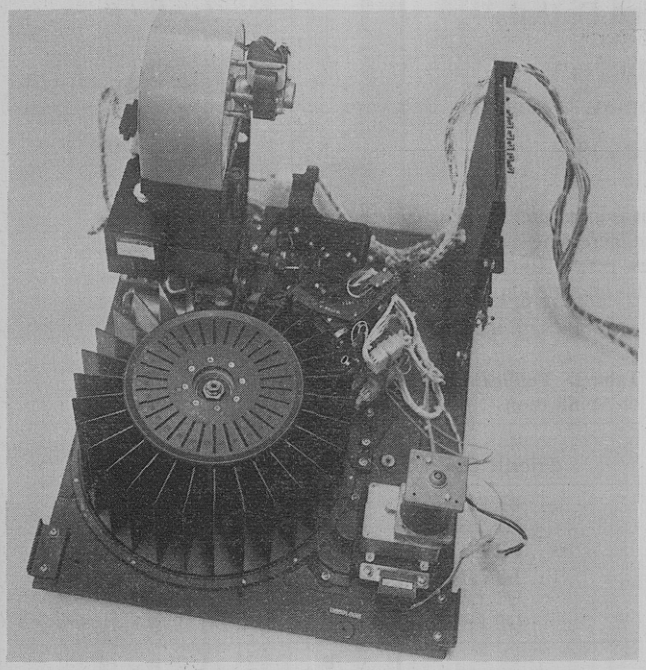

Figure 3. Carousel and display mechanism.

explanatory information relating to execution of the INSTRUCTION to which it is related and which can be accessed by pressing a large key labeled "HELP". If a user is unfamiliar with the CCMSS, it is recommended that he or she make frequent use of the HELP messages; as the searcher becomes more accustomed to using the system, there is less need for referring to them.

At this time the CCMSS contains indexed files in three art areas: General Purpose Digital Processing Systems, Special Purpose Digital Processing Systems, and Analog-Digital Convertors. Three Questicon units, each containing microfiche for all three files, are located in the Examining Corps and one in the Office of Search Systems for use by members of the public. Files are updated quarterly by the contractor.

One of the large classified files selected for implementation under the new contract is in the field of Synthetic Resins, Classes 526 and 528. These two classes together now contain about 20 000 discrete U.S. patent documents classified and cross-referenced in paper file containing about 68 000 paper copies. Availability of the file on the CCMSS is scheduled to occur in July of this year. At that time four Questicon units will be available for use by Examiners in Group 140. In addition a Public Search Center will be established, on the second floor of Building 3 in Crystal Plaza, where a unit will be available for use by the public.

One advantage which patent examiners see the CCMSS as offering, to offset the disadvantages of using microforms, is the EXCLUDE feature. Searching the paper file frequently involves searching several related subclasses; the same document may be located in two or more of these subclasses. Searching the CCMSS will enable examiners and other searchers to avoid looking again at any document they have already viewed since the beginning of a terminal session.

One other interesting feature of the software is "tagging". A searcher browsing through documents can press a key labeled TAG/UNTAG while viewing ones of particular interest. At any point the searcher can interrupt browsing to go back and review those previously tagged. Documents can be removed from the "tagged list" if they are no longer of interest (e.g., because a more pertinent reference was found later) by pressing the same key. At the end of the search session it will be possible to obtain a printed search report containing the subclasses searched, the number of documents viewed, and a list of the documents which have been tagged.

The impact of a computer system such as the CCMSS on examination time and pendency is expected to be small; however, improvement in quality of search, which will necessarily follow from the file integrity inherent in such a system and so sorely lacking in the present paper system, is expected to be very substantial and contribute significantly to reducing and eliminating the sharp criticism presently being leveled at the Patent and Trademark Office regarding lack of file integrity.

In addition to the examiner time which will be saved, there will be elimination of the clerical time now required for refiling documents which examiners have removed during their work.

Dollar savings can also be expected from the savings of space which will occur in the Examining Corps and in the Public Search Room if such systems prove feasible and their use becomes widespread.

Computer systems such as the CCMSS can have terminals located in remote locations and thus offer the opportunity to improve service to the public by making feasible the establishment of one or more search centers in places outside the Patent and Trademark Office in Arlington, Va.

Upon installation and evaluation of the proposed systems for searching classified files, if warranted, an effort will be made to obtain a software package which will enable reclassification of patents on-line at a terminal by a classifier using much the same thought processes and creative techniques now used in creating a new classification search system.

Substantial benefits could flow both to the office and to the public if such on-line capability is developed. The clerical processing to build the new paper file following the professional effort by the classifiers and technicians costs in excess of $1.50 per patent reclassified, and hundreds of thousands of patents are reclassified each year. Further, with the present procedure, large numbers of patents are unavailable in the Public Search Room for periods up to two years while they are being reclassified and the new file is being prepared.

If classifiers are able to work on-line at terminals, this substantial clerical effort required to build new paper files for the examiners and the public can be eliminated. Further, there need be no period during which files would be unavailable, other than the time required to read in a data tape and insert a few pieces of microfilm into each terminal.

For the future, it is anticipated that classified search files which are implemented on the CCMSS may have supplemental index terms such as from IFI applied to portions of the art. Further, International Patent Classification symbols can also be stored in the system and used for searching. Both of these additional ways of accessing documents can be employed at little cost and will significantly decrease search time in many arts while contributing to improved quality; neither is practical for classified files implemented with paper.

REFERENCES AND NOTES

(1) 35 USC 101-3, 112.
(2) U.S. Department of Commerce, "Development and Use of Patent Classification Systems", U.S. Government Printing Office, Washington, D.C., 1966.
(3) P. Glickert, "Patent Office Classification: Its Whats and Whys", *J. Am. Soc. Inf. Sci.*, **25**, 308–311 (1974).
(4) S. M. Newman, "An Operative Information Retrieval System Based on Relational Factors", *Inf. Storage Retr.*, **3**, 331–350 (1967).
(5) C. V. Clark, "Obsolescence of the Patent Literature", *J. Doc.*, **32**, 32–52 (1967).
(6) P. M. McDonnell, "Technical Information Management in the U.S. Patent Office", *J. Chem. Doc.*, **9**, 220–4 (1969).
(7) P. M. McDonnell, "ICIREPAT and International Developments in Patent Information Retrieval", *Spec. Libr.*, **66**, 133–9 (1975).
(8) P. M. McDonnell, "A New Approach to Mechanized Information Retrieval in the U.S. Patent Office", *J. Pat. Off. Soc.*, **50**, 651–8 (1968).
(9) Image Systems, Inc., P.O. Box 2488, 11244 Playa Ct., Culver City, Calif. 90230.