

Patent Searching in a Pharmaceutical Company[†]

TRISHA M. JOHNS* and DOROTHY I. RYNO

G. D. Searle & Company, Box 1045, Skokie, Illinois 60076

Received September 15, 1977

The patent searching activity at G. D. Searle and Company is described.

The Information Services Department is set up to serve the needs of the world-wide Searle organization. The Department is responsible for storage, retrieval, and dissemination of information both internally generated and publicly available. The Department is composed of two major areas: unpublished information including internal data, reports, and information on the health-care industry; and published information including the scientific literature and patents. Requests to the Department are submitted on standard forms, through phone calls, letters, or telexes from non-Skokie locations. All requests come to the attention of all Department supervisors to ensure that those of a multidisciplinary nature are handled by information scientists expert in those areas. One person is responsible for collecting, collating, and responding to the user for any given search.

The Chemical Literature and Patents Section is a part of the published literature function of the Department, and its primary responsibility is to answer on-demand search requests. Roughly one-third of all requests submitted to this group involve a patent search. (In 1976, this meant about 300 patent searches.) There are six information scientists in the Chemical Literature and Patents Section, each with a degree in chemistry. Those aspects of information science necessary for chemical information work, such as nomenclature, Wiswesser Line Notation, chemical fragment codes, and use of on-line services, are learned on-the-job. Advantage is taken of seminars, in-house training courses, and meetings both for the new and the experienced information scientists. Requests to this Section come chiefly from the Chemical Research Department of Searle Laboratories, the Corporate Patent Department, and Searle's research facility in England.

At Searle Laboratories, interdisciplinary committees including chemists, biologists, and clinicians (i.e., M.D.s) determine the general kinds of research to be conducted. The chemist, working within his group, formulates a set of compounds which may be expected to be active against specific disease states. Although the chemist is usually well aware of work being done in his field, he will often ask for a general "background" literature search. Before beginning laboratory work, the chemist may request that a search of patent literature be done as well, to ensure against the proverbial "reinventing of someone else's wheel". After the compounds have been synthesized and are sufficiently identified and purified, they are ready for the initial animal testing.

In addition to synthesis and submission of compounds to appropriate animal screening tests, the chemist is also responsible for initiating patent action to protect Searle's interest in the new compounds. The Corporate Patent Department submits another search when preparing the patent application for submission to the U.S. Patent and Trademark Office. It is not uncommon for the same, or very similar, search to be submitted more than once. Searches are knowingly repeated as a double-check.

The most common type of request is to determine novelty of a compound or a series of compounds. It may be necessary

to search for compounds in conjunction with specific uses, or novelty of processes for preparing, manufacturing, extracting, resolving, or purifying compounds. Often, the requestor wants to know about patents issued to certain individuals or companies. A list of foreign counterparts is often desired, as are copies of U.S. and foreign specifications.

The sources of patent information used by the Section are: Derwent Publication's *Central Patents Index*, IFI Plenum Data Company's CLAIMS files, *Chemical Abstracts*, *Official Gazette* of the U.S. Patent and Trademark Office, and an internally generated file called the "External File". The type of request determines which database is used, depending on the content and available search parameters.

Derwent's *Central Patents Index* (CPI), Farmdoc Section, is a major source of information on world-wide pharmaceutical patents. Coverage in the major countries goes back to 1964. The Derwent Farmdoc Section includes pharmaceutical and veterinary specifications, intermediates for use in the manufacture of drugs, the production of pills, capsules, and the like, compositions used for diagnosis and analysis, and devices for dispensing pharmaceuticals (syringes, child-proof bottles, calendar-pill boxes, aerosols).

For compound novelty, searches of chemical structures and substructures are carried out through the Derwent Chemical Punch Code. This is a fragment code, based on punch positions on an 80-column punch card, each card-punch representing a different chemical fragment. The important thing to keep in mind when searching with fragment codes is that the compound will drop with any subset of punches describing it. Therefore, in a search for ascorbic acid, the patent would appear if one of these punches or any number of them were searched. The position to strive for, of course, is the ability to use only punches that fit the compound and at the same time reduce irrelevant citations. The use of a fragment code permits coding of Markush structures, but this same feature that allows for broad compound description results in many false drops. In the earlier patents, Derwent coded all the compounds from a patent on *one* punch card, and this so-called overcoding only added to the inherent problem with Markush structures. In recent years, however, Derwent has done much to alleviate this problem by sorting the punches for each compound separately in a given patent. Collaboration among information scientists is sometimes necessary in interpretation of the coding rules, and even the experienced information scientist benefits from the Derwent-sponsored coding classes.

In 1972, the Department developed an on-line computer system to handle the Derwent Farmdoc data. For the past five years this system has been maintained, as until recently when System Development Corporation took up the banner, there was no alternative to batch searching the Derwent tapes. Our on-line system allows for on-the-spot query modification and, because it is run in-house, for quick delivery of off-line prints (within hours of the search).

The CPI information, 1964 to present, which includes the bibliographic citations, dictionary of searchable terms (the so-called manual codes and assignees), and the punch codes are stored on two removable disks. These disk packs are "called up" on-demand. The information scientist inputs a

[†] Presented in the symposium, "Trends in Handling Patent Information", before the Division of Chemical Information, 174th National Meeting of the American Chemical Society, Chicago, Ill., Aug 29, 1977.

message to mount the packs via a computer terminal, and in most cases the operator is able to put them up within five minutes. The information scientist enters the search strategy, having the option to list the abstract numbers, get full citations printed on- or off-line, or modify the strategy.

Fragment codes can be ambiguous and the bulk of CPI searches result in many potentially relevant citations. The information scientist must scan each citation and, on the basis of title, choose those hits that appear valid. The result is a subset of hits, the abstracts of which must be consulted. These are found on the punched cards supplied by Derwent weekly. The cards are filed in accession number order. If the abstract contains insufficient information, the complete specification must be read.

The Derwent titles of older specifications are often non-informative, for example, "Heterocyclic Compounds" or "Steroids", which means that the abstract cards must be pulled, checked, and refiled every time that number is a hit on a search. In an attempt to make this step of the search more efficient, key compounds of the older Farmdoc patents are being generically coded into Wiswesser Line Notation (WLN). This notation will be added as another title line, as a visual aid to determining relevancy of a patent.

The WLN will not be used for searching as it was not the intent of the project to code all the compounds covered in the patent. WLN is best suited to describe discrete molecules, and it is difficult to generically code compounds as they are described in the patent literature.

The Derwent patents database is available on-line to Derwent subscribers through System Development Corporation (SDC). Although SDC's Orbit language differs from that of our internal system, the approach to the search is the same. SDC's PATENTS database is accessed on occasion because a greater number of searchable parameters have been made available (e.g., title terms, application numbers). Also there are times when the search is nonpharmaceutical and through SDC the whole of CPI is available.

IFI Plenum Data Company's CLAIMS files available through Lockheed's Dialog Retrieval Service is used primarily for patent searches that do not specifically define a chemical structure. CLAIMS covers over 360 000 U.S. chemical patents and chemically related patents from 1950 to the present, including foreign equivalents from selected countries. CLAIMS is searchable through three types of terms: general title terms, U.S. Patent Office Classification Codes, and assignees. The list of classes and subclasses can be found in the U.S. Patent Office *Manual of Classification*, published by the Government Printing Office. Often appropriate class codes are determined by entering pertinent keywords. The class codes found on the "hits" are then used in a further search.

When searching for assignees in CLAIMS, an assignee code number is used, which circumvents the problems of variable

names for one company. These code numbers can be found in the listing from IFI Plenum Data Company called *Frequency Distribution of Patents by Assignee*.

Chemical Abstracts is the starting point for any search for chemical information. Searching for compound novelty demands a thorough look-up in both subject and formula indexes. CA selects chemical patents from 26 countries and, like Derwent, publishes the first occurrence of a given invention and draws up lists of equivalents. In addition, CA and its associated computer-retrievable databases are convenient places to search for individual inventors.

Another source drawn upon by this Section is the so-called External File. This is an internally generated and maintained WLN file of over 300 000 compounds. The compounds have been coded and accumulated for over two decades, and include a large holding of compounds from U.S. steroid patents, selected pre-Farmdoc Derwent patents, basic prostaglandin patents, and others deemed, now or in the past, to be of major interest to Searle.

When answering search requests, the original document must often be consulted, and, as a result, it has been necessary to subscribe to the complete specifications of all Derwent Farmdoc basic patents and U.S. chemical patents on microfilm. An outgrowth of these holdings has been a Patent Copy Service. Copies of U.S. and foreign patents are supplied on-demand. Often the requestor prefers an English-language equivalent patent, and the clerical staff has learned to use the Derwent and CA patent concordances. If the U.S. patent required is not held in-house (i.e., if it predates the collection, or if it is nonchemical), it is ordered from the U.S. Patent Office. For rush requests, the Chicago Public Library's complete collection of U.S. and British patents is available. Derwent foreign equivalents are obtained through their Patent Supply Service.

To keep the research staff up-to-date, 70 copies of Derwent's CPI Farmdoc Alerting Bulletin are distributed. These are booklets of approximately 50 pages, received weekly, and classified according to subject matter. They contain abstracts and pertinent documentation to patents selected as basic and including most foreign equivalents. The research investigator can, in a reasonable span of time, stay aware of patents in his specialty. For a number of years, bulletins of U.S. patents "of interest" from the weekly Official Gazette were compiled and distributed. It was discovered that Derwent covered the majority of the patents in the CPI Alerting Bulletin. Switching over to these low-cost current-awareness books saved considerable time and effort for the staff.

The development of a drug from its conception to the day it takes its place in the PDR, is a time-consuming and costly venture, and the years of patent protection for a drug are short. Patent information services not only assist in securing that protection but, just as importantly, play a fundamental role in the direction of the R&D effort.