PATENTS, LITERATURE

1. Introduction

Patent systems were established during the Industrial Revolution of the eighteenth and nineteenth centuries to provide incentive for the development of technology by later inventors and to make information available about the advances made by previous innovators. Patent rights are limited to control of the manufacture and use and sale of the invention claimed in the patent; a patentee has the right to practice the claimed invention only to the extent that this does not require practicing inventions claimed in patents owned by others. The patentee has the right to license, assign, or sell the rights conferred by the patent and to sue for infringement, unauthorized manufacture, use, or sale of a product, a process, or an apparatus covered by the patent claims. Patents are granted by national governments and have effect only within the granting state. Patent rights and collections of patent literature documenting these rights existed in each industrialized country, but copies of patent documents were not widely available throughout the world. The internationalization of commerce in the late nineteenth century necessitated the filing of patent applications in each country where the innovator wished to exclude others from practicing the invention, providing motivation for the founding of an international treaty establishing standardized treatment of patent applicants. Since 1883, the Paris Convention for the Protection of Industrial Property has guaranteed that an applicant for a patent in one member state may file applications for patents in all other member states within 1 year of the original filing date and will be given rights to the claimed invention as of the priority date established by the first filing.

The second half of the twentieth century has witnessed a sharp increase of activity in research and development, as well as an increased internationalization of technology-based industries. As a result, there have been significant changes in the patent literature, the chief literature of technology. The number of countries publishing patent documents has increased as former communist and Third World countries have enacted patent laws. The end of the twentieth century saw growth in the number of patent-issuing authorities as a result of the emergence of new nations formerly embedded within the Soviet Union and other communist countries of central and eastern Europe, and of the enactment of new patent laws by other countries in response to the intellectual property provisions of the General Agreement on Tariffs and Trade (GATT), which established the World Trade Organization (WTO), and to the North American Free Trade Agreement (NAFTA). The ideal of full harmonization of patent laws among countries has often been discussed, but seems far from being realized; nevertheless, significant changes have been made in the patent laws and procedures of individual countries.

The changes have been in patent documents themselves as well as in the means of documentation. The need to cope with a rapidly increasing volume of patent applications led a number of patent offices to switch, mostly between

1964 and 1979, from a system in which all applications were examined and only those found to be worthy were issued a patent, to a system in which all applications are published and may or may not be examined at a later date or will ever become a patent. Pregrant publication became virtually universal with the introduction of published patent applications by the United States in 2001. In the 1940s, most published patent documents were patents; by the 1990s most were unexamined applications.

Until the 1970s, all patents were effective only for the individual countries by which they were issued. In the twenty-first century there are several types of international patent documents, eg, the European patent, which provides patent rights granted centrally by the European Patent Organization (EPO), and which is enforceable after national registration in as many as 30 European countries at this writing. A Eurasian Patent Convention was established in 1994 by 11 former member states of the Soviet Union. Two regional organizations cover a number of African nations: the African Intellectual Property Organization (OAPI) for 14 French-speaking countries and the African Regional Industrial Property Organization (ARIPO) for 11 English-speaking countries. The Gulf Cooperation Council grants patents to 6 countries in the Middle East. Patent Cooperation Treaty (PCT) applications provide a means for filing applications for patents in multiple patenting authorities via a single application. PCT applications receive a preliminary examination and are published by the World Intellectual Property Organization (WIPO) prior to examination under the national patent laws of each of the member states (128 in 2005) designated by the applicant. Rights under granted European patents can at present be extended to six additional PCT contracting states,; it is expected that those countries will become full members of the EPO and it is possible that additional PCT states will be added to this list. Negotiations within the European Union to create a Community Patent, a true European patent granted and enforceable by a single agency, have been revived repeatedly, but have always failed over issues of language and sovereignty.

The development of computerized publication systems in the 1970s allowed patent offices to begin publishing patent documents in electronic format, and the later development of secure Internet connections has allowed patent offices to begin accepting applications, storing application files and corresponding with applicants electronically rather than on paper. This allows twenty-first century patent documents to be distributed globally on the day of publication and facilitates the sharing of detailed information about the status of published applications and patents. Electronic publishing and distribution of patents has expanded access to patent literature from a limited community of specialists to anyone in the world with interest and a high speed Internet connection. Table 1 shows some of the milestones in the development of the primary patent literature during the last third of the twentieth and the early years of the twenty-first century.

Patent laws provide for several stages in the life of an application for a patent on an invention. The pattern followed by patent laws in effect in most industrialized countries during the nineteenth and early twentieth centuries, and still in effect for some applications filed in the United States in 2005, calls for the examination of all patent applications to certify that the claimed invention meets the national standards for novelty, usefulness, and inventiveness. The

owner of the technology to be patented files application papers that include a specification containing a description of the invention to be patented (called the disclosure) and claims defining the limits of the invention to be protected by the patent, a formal request for the issuance of a patent, and fees. Drawings of devices and apparatuses, electrical circuits, flow charts, etc, are an important part of the disclosures of most nonchemical and many chemical patents.

The national patent-issuing authority assigns an application serial number, examines the application papers to see that all requirements are met, examines the claims of the application to determine whether patent rights are justified in view of the earlier disclosures in the prior art, and corresponds with the applicant to negotiate any amendments that might be required. When a suitable scope for the claimed invention is agreed to, the patent is granted and the patent specification is published. When the patent examiner determines that no patentable invention has been claimed, the patent application is abandoned. Under this scheme, unexamined applications are not published and the patent office provides no direct evidence that such applications had been filed. However, indirect evidence could sometimes be obtained from related patents issued to the patentee or assignee. The U.S. law with regard to the publication of pending patent applications changed with effect in 2001, and the only patents that are published for the first time on their grant date are those whose owners certify that no corresponding application is being filed in a country with 18-month publication.

Under U.S. law, the inventor is defined as the owner of the patent unless the patent rights have been assigned to his or her employer, or some other individual or organization. Designations of assignment are typically filed with the U.S. Patent and Trademark Office (U.S. PTO) prior to the issuance of patents, and the name of the assignee is printed on the patent. In most countries outside of the United States, the patentee is the employer, rather than the employed inventor.

As the number of patent applications filed during the middle of the twentieth century grew, the time required to notify the public that an invention had been claimed in a pending patent application was seen as a serious inconvenience. Laws were introduced in some countries to inform the public about potential patents during their pendency. In some countries, the name of the applicant, the title of the patent application, and the serial number are published immediately after the filing of the patent application. The full patent specification is published or made "open for public inspection" (OPI) by most modern patent offices approximately 18 months after the original filing date of the application. The first publication of the patent application usually follows some initial examination of the application by patent examiners, but no judgment as to the patentability of the claimed invention is made then and the published specification will not ordinarily have been amended. Full examination of the patent application is omitted by some countries unless the validity of the claims is challenged by a third party. Most countries, however, proceed with the examination and publish the amended specification for a second time when the patent is eventually granted. The validity of the allowed patent claims can be challenged in formal opposition proceedings after the formal grant of the patent. If the opposition

proceedings result in modification of the patent claims, an amended granted patent may be published.

Once granted, patents are in force for a term prescribed by law. Patent terms are not renewable. Most countries have established a term of 20 years, measured from their national filing date, but patent laws enacted before the latter third of the twentieth century vary considerably in the length of the patent term. Japanese patents have a term of 20 years from filing, but until 1995, the term was subject to the limitation that it would expire no later than 15 years from the date of grant. Under U.S. law, patents based on applications made before June 8, 1995 have an effective term of 17 years from their grant date, regardless of the length of the interval between filing and issue. Patents filed on or after June 8, 1995 have a term of 20 years from their original U.S. filing date. Those patents granted under the earlier law and still in force on June 8, 1995 have been given a revised expiration date of 20 years after their earliest U.S. filing if the regular 17-year term would cause them to expire earlier than that. The new U.S. law also permits the filing of incomplete provisional applications, which do not require proposed claims, as domestic priorities. A regular application must be made within a year from the filing of a provisional application. The domestic priority period will not count in the life of an issued patent derived from a provisional application. Patents issued by the United Kingdom before the current 20-year term was established had a term of 16 years from filing. Terms of 16 and 17 years were once common, but have been supplanted by 20-year terms in nearly all countries.

The expiration date of a patent is not normally printed on its face and must be calculated on the basis of the applicable national laws. Exceptions can occur in the United States when a term is foreshortened because of that patent's close relationship to a previously issued patent or, under the new law, when a term is extended because of delays in the course of patent prosecution. In addition, patents issued by most countries are kept in force by payment of periodic maintenance fees. Because some products cannot be marketed without the approval of governmental regulatory agencies, the owners of patents on drugs, medical devices, and agricultural chemicals have long complained that the effective term of their patents is less than the term of unregulated products. Some countries have provisions for the extension of the patent term for products approved for marketing under regulatory laws. Patent term extensions granted by the United States and Supplementary Protection Certificates granted by EPO member states are effective only for the approved product and not for other products that might be covered by the patent. To determine whether a particular patent is in force, it is necessary to obtain information about the current legal status of the patent from sources other than the patent specification itself. A compilation of national and international laws regarding patent expiration has been published by Thomson Derwent (1).

1.1. Patenting Procedures. Procedural pathways followed by patent applications filed in various countries and resulting in the publication of patent documents are shown in Fig. 1. National filing of a patent application in the home country of the applicant is typically treated as priority filing under the Paris Convention for the Protection of Industrial Property. Patent applications may be filed directly in each country of interest to the applicant, or a single

application under the Patent Cooperation Treaty may be filed to facilitate acquisition of patent rights in many countries. If the PCT is chosen, a preliminary examination is performed and the patent specification is published with a search report prior to the applicant's filing in each designated patent office, where national procedures are followed. Countries, eg, South Africa, follow the simplest procedures, examining the application for formal compliance with the patent laws and publishing the specification as a granted patent without examining the claims for novelty and nonobviousness. In such countries, the validity of the patent is tested only when the patent is challenged in the courts by a third party. Long-standing U.S. practice, now applicable only to patent applications without counterparts in countries with 18-month publication procedures, withholds publication until all aspects of examination have been completed and the patent rights granted. No opposition by third parties is provided for in the United States, although on rare occasions a patent may be reissued to correct irregularities in patent prosecution or reexamined to permit reconsideration of prior art overlooked before the grant of the patent.

Most patents follow a more complex procedure. The patent specification is published before a substantive examination of the patent application. In Japan and the United States, publication takes place before a search of the prior art has been performed. In the EPO and in preliminary examination of PCT applications, a search is made before the specification is published, and the search report is published at the same time or shortly afterward to assist the applicant in deciding whether to continue prosecution. If the claims are determined to be patentable by the patent office, the patent application is granted and opened to opposition by third parties. An amended patent will be issued if the scope of the patent is changed during opposition proceedings. In Japan prior to 1996, the opposition period took place before patent rights were granted; beginning in 1996, examination is followed directly by grant, as in the EPO, and the opposition period takes place after grant. Legislative and procedural changes take place from time to time; most countries that established or revised patent laws during the late twentieth and early twenty-first century have adopted laws providing for terms of 20 years with publication 18 months after filing. For example, Canadian patent law paralleled that of the United States until 1989, when the former adopted a new patent law based on EPO procedures.

1.2. Patent Documents. The internal structure of patent documents has been standardized and the amount of bibliographic detail recorded in a patent document has increased. Early patent documents included rudimentary information about the patent's filing details. Most patent documents published during the twenty-first century begin with an informative cover page; an example is shown in Fig. 2. The front page of a modern patent provides key information about the patent that aids the reader greatly in determining the patent's potential relevance. The cover page provides a title, gives the name of the patent owner, inventors, and other individuals involved in the issuance of the patent, and offers serial numbers and dates that identify the document and relate it to other patent documents covering the same invention. An abstract is provided by the patentee. Where appropriate, the abstract may include structural diagrams for chemical species important to the invention, or a representative drawing. National and International Patent Classification appropriate to the patent are

shown, as is the list of classes searched by the examiner in determining patentability. Patent and other publications deemed by the examiner to be related to the invention are listed, and are referred to as examiner's citations. These bibliographic data have been standardized according to Internationally agreed Numbers for Identification of Data (INID) codes established by WIPO. The INID codes provide a means whereby the various data appearing on the first page of a patent and other similar documents can be identified without knowledge of the language used and the laws applied. They are used by most patent offices and have been applied to U.S. patents since August 4, 1970 (2).

Because each country has its own patent laws, the precise meaning of the bibliographic data and the legal significance of the published patent document vary from country to country. The PCT provides a recommended code to distinguish the various types of documents and to simplify storage and retrieval of patent data (3), but the code is implemented differently by different countries. For example, in the United States an A-document published before 2001 was a patent; in the Netherlands, an A-document was a published unexamined application. When the United States introduced pregrant publication of patent applications in 2001, those were A1 publications, and granted patents without a previous publication were identified by the code B1. It is essential to understand each country's system to interpret the status of its patent documents.

The invention covered by the patent is defined in the patent claims, which appear at the end of patents of most countries and at the beginning of patents published by a few others. The majority of patents are known as utility patents. The claims of these patents may relate to new products, including new chemical compounds and compositions, to processes for making or using new or previously known products, and to machines for making products or using processes. Patent claims are examined by the national or international patenting authority to determine whether the claimed subject matter is patentable under the applicable statutes and by comparison with the prior art in the field of the invention. Where the claims are judged to cover more than one invention, the patent examiner may restrict the claims to a single inventive entity and authorize the filing of divisional patent applications. Where the inventor wishes to modify the invention, continuation-in-part applications or applications for patents of addition may be filed. What may be claimed in a patent differs from country to country, and has changed over time with amendments of national patent laws. Increasingly, patents are granted only when the claimed invention is novel, ie, has never before been patented, practiced, or described in a published document anywhere in the world; possesses an inventive step or is not obvious in light of the prior art; and has utility for a purpose acceptable under the law. Novelty and nonobviousness are determined by performing a search of the published patent and nonpatent literature. Many countries have limited patent rights to inventions considered socially useful or having industrial applicability. Methods of treating the human body, foods, pharmaceutical compositions, chemical compounds per se, living organisms, atomic weapons, computer programs, and scientific theories have all been held to be unpatentable in many countries at various times. Although the claims of granted patents are strictly limited by the various national laws, patent applications published before grant often claim subject

matter that is not patentable; consequently, the claims of granted patents often differ significantly from the claims published in the unexamined application.

In addition to utility patents, some countries publish patent documents under different or less stringent standards for patentability and with shorter patent terms. For example, U.S. plant patents cover asexually reproduced plants. United States design patents cover the decorative aspects of a product, a function served in most other countries through industrial design registration systems. Utility models and petty patents cover products with differences from the prior art that need not meet the nonobviousness standards set for utility patents.

The bulk of the patent specification is the disclosure, the text and illustrations that describe the claimed invention in detail and explain how the claimed invention differs from the prior art. Modern patent disclosures contain a summary of the claimed invention, a description of the background of the invention, a general description of the way in which the invention is made and used, specific examples, and, where applicable, drawings of the invention in general or specific embodiments. The technical information provided in a patent specification may be used without infringing the patent; only practicing the invention defined in the claims within the term and territory of the patent grant is forbidden. Because much of the information in patent specifications is never published in refereed journals or other nonpatent media, patent disclosures are an invaluable part of the technical literature.

Patent documents differ from journal literature in several ways. First of all, they are legal documents whose disclosures support one or more claims that define an area of property rights. The language in patent documents can therefore be quite convoluted "patentese" as the applicant strives to achieve the broadest possible scope of coverage. Examples provided in patents may never have happened. Based on the applicant's understanding of the technical area involved, he or she may assume the probable outcome of experiments never actually run, and include such paper examples in the patent specification. Paper examples are generally written in the present tense. They lack hard data, and can provide grounds for attacking the patent should they prove to be inoperable. Finally, chemical patent disclosures and claims can be written in terms of generic structures, or the so-called Markush structures, in which one or more portions of a chemical entity can vary, including functional groups, numbers of substituents, and points of attachment. Markush structures are used as one method of obtaining the broadest possible claims in a patent, and are named after an early inventor who succeeded in obtaining claims on a process for making such variable products. Markush structures can be simple, describing just a handful of chemical compounds, or highly complex, encompassing thousands, millions, even infinite numbers of compounds; a typical example (4) is shown in Fig. 3. The effective indexing and searching of Markush structures provides a significant challenge to those concerned with chemical patents (5,6).

1.3. Defensive Publications. Published patent applications form the bulk of the patent literature, and a great many of these applications are rejected during examination or are voluntarily abandoned and never continue through the examination process to become granted patents. These documents serve as defensive publications: they introduce inventions to the prior art and block

others from obtaining a patent, ensuring the applicants that they will have freedom to practice their inventions without fear that a competitor will obtain a patent in the future. Prior to the introduction of pregrant publication of patent applications, the U. S. Patent and Trademark Office published a series of Defensive Publicationsm and later Statutory Invention Registrations (SIRs), often applied for in lieu of a patent on an invention that had been rejected during examination.

It is also possible to protect an innovation from the possibility of infringing future patents by publishing an invention disclosure in any form. There are defensive disclosure publications designed specifically as a quick and inexpensive format for publishing invention disclosures. For many years, the print publication *Research Disclosures* has served in this way. Derwent and Chemical Abstracts include *Research Disclosures* in their patent coverage under the quasicountry code RD. Derwent began coverage in 1978 and discontinued it in 2001, while CAS began coverage in 1996 and continues to cover *Research Disclosures* in 2005. An additional printed service, *International Technology Disclosures*, was covered by Derwent from 1984 to 1993. For many years, IBM published its own inventions and those of others in the. A new electronic defensive disclosure service, IP.com, was established in 2001, and is indexed by CAS and accessible from the Delphion Intellectual Property Network as well as its own website. IP.com serves as a forum for technology transfer as well as a defensive publication site. Another technology transfer service, Yet2.com, also permits users to post descriptions of technologies that are not covered by patents. These defensive disclosures need not be alternatives to patents; posting a defensive disclosure in an online service can give an early notification to potential licensees and competitors while the owner of the technology waits for the 18-month publication of a patent application.

1.4. Patent Families. Patent specifications are published as individual documents in the language of the originating country, but many inventions are claimed in patents issued by more than one country. These patents form a family of equivalent patent documents, which usually disclose the same information but may differ somewhat in the scope of their claims. When filing in more than one country, an applicant establishing priority under the Paris Convention is generally required to submit a copy of the original application to each national or regional patent office selected, sometimes with a translation. A simple patent family is based on a single priority application, in which each family member discloses the same information and cites the same priority application number. When the technical content of the patent rather than its legal scope is of interest, any member of the family can be substituted for another, thus often obviating the need for translation.

When an applicant misses the deadline for convention-filing or files applications in countries that are not members of the Paris Convention, the priority application number will not be present in all patent applications having equivalent disclosures and claims. Nonconvention equivalents, which can only be recognized by comparing the contents of the applications, form what WIPO defines as an artificial, intellectual, technical, or nonconvention family. When the applicant has refiled the original national patent application within the priority year, more than one priority application may be claimed in the corresponding foreign

applications. Applicants are also permitted to combine the disclosures of two or more patent applications within the priority year and file foreign counterparts incorporating information from each. A complex patent family contains patent documents having at least one common originating application. An extended patent family contains all of the patent documents having at least one priority application in common with any other member of the extended family. Extended patent families sometimes include patent documents that differ radically from other family members. Members of a national patent family, one that includes divisional, continuation, continuation-in-part, and addition patents issued in a single country, may also differ significantly in content. Patent databases usually provide technical information from a representative member of a patent family, which defines the patent family according to one of the definitions above or devises a hybrid definition unique to the database.

1.5. Patent Searches. Because valid patent claims can only be issued on an invention that is novel and innovative in light of prior art, it is necessary to search the prior art for previous references either to the composition of matter, process, or machine defined in the claims of a patent application, or to any similar composition, process, or apparatus that would render the claimed invention obvious to a person skilled in the field of the invention. Inventions that have been described in a publication or embodied in a product are said to have been anticipated in the prior art and are not patentable. Patentability searches are performed by examiners employed by the national and regional patent offices and are an important step in the examination of patent applications. Patentability searches should also be performed by the representatives of inventors prior to the filing of a patent application so that the claims will not overlap with any publication in the prior art. These searches may encompass the full scope of the published literature, including patents, technical journals, gray literature, and even catalogs. Individuals or organizations who are making plans to introduce a new product or process must conduct infringement searches to ensure that they will not infringe patents that belong to others. Searches whose objective is avoiding infringement can be designated as freedom to practice, freedom to use or freedom to operate; they need only consider patents in force and pending applications that may result in patents in countries where manufacturing or marketing are contemplated. After a patent application has been published and/or a patent has been granted, organizations that wish to practice the invention may also conduct validity searches to be used as ammunition for opposition proceedings or invalidity lawsuits. Validity searches, like patentability searches, should include all forms of published literature, but are limited to publications with effective dates earlier than the filing date of the patent application being challenged.

Searches of scientific and technical literature are performed using any of the information retrieval tools suitable for searches done for other purposes (see INFORMATION RETRIEVAL). During the nineteenth and twentieth centuries, the primary source of patent information was printed patent specifications, which were searchable only through patent classification codes. Access to printed patent specifications was available at the patent offices, which maintained public search facilities where classified files of patents could be searched. Patent offices have now migrated to electronic publication of patents; bibliographic information, full patent text, and/or document images can be viewed anywhere in the

world on the date of publication. Patent offices have devised special classification systems to facilitate searches among the individual patent documents in their collections. These patent classification systems were designed to subdivide patents into groups covering similar inventions that could be reviewed by examiners when related inventions were claimed in later applications. All of the existing fields of science and technology were defined and provided with a class code and subdivisions of the fields were given narrower classification designations. Patents belonging to each subclass were originally stacked together in drawers or on shelves similar to the stacks of boxes in a shoestore, and examiners or members of the public could extract a stack of patents and search for information in the subfield of interest by flipping through paper copies of the patent documents. As new fields of science and technology have developed, each patent classification system has been revised so that the emerging technologies can be searched. Patents are assigned classification codes by the examining office and the relevant primary classification and any cross-reference classifications are printed on the first page of the patent, eg, INID codes [51] and [52] in Fig. 2. Although patent classifications originated as tools for manual searches, they can be searched through printed or electronic indexes as well, and electronic search systems have become the norm by 2005.

National patent offices created patent classification systems for internal use without correlating their guidelines for subdividing technologies or the symbols used to identify classifications with those of other countries. The assignment of national classification codes to patents of the issuing country facilitates searching for inventions claimed in national patents, but is not helpful for prior art searches that must include patents issued by patent offices that use different classification systems. Therefore, the internationalization of commerce has led to the internationalization of patent classifications. The first edition of the International Patent Classification (IPC) system was introduced in 1968; it has been revised by WIPO every five years and was in its seventh edition in 2005 (7). A "reformed" IPC system will go into effect in 2006. At that time the update frequency for the core classification will be reduced to 3 years and an advanced level of classification will be updated on a 3-month cycle with updated classification codes published electronically.

The IPC has been adopted by most of the patent-issuing countries of the world. Even countries such as the United States, that continue to use a national classification system to organize their patent search files, print a corresponding IPC classification on the patent documents. Although the IPC is used by most countries, these countries do not all follow the same guidelines for applying the codes, nor do they all use the finest divisions of the classification system. Differences in the scope of the patent claims, on which the IPC classification is based, as well as differences in the classifying examiner's interpretation of the novel features of the invention, also contribute to differences in patent classification among countries. It is not unusual for patents having identical claims to be classified differently in each country where the patent application was filed.

The IPC codes, which have the format *ANNA NNN/NN*, where *A* stands for a letter and *N* a numeral, represent a hierarchical system (7). The first four characters designate the section, class, and subclass of the class code, and each successive character narrows the definition of the invention. Each subclass

is further divided into groups, defined by one to three numerals and followed by two- to three-digit subgroup designations. The hierarchical relationships within the groups are determined by the relationships published in the IPC manual (Fig. 4). Historically, some countries have indexed patents only to the four-character subclass level (termed the core level under IPC 8), whereas most others have used the full IPC (the advanced level). Although a main IPC code is always provided, some countries assign supplementary and/or additional IPC codes that designate additional aspects of the claimed invention. In searching, it is usually necessary to truncate IPC codes unless the level of specificity used by the country of interest is known.

The accellerated development of new technologies made the rigid IPC schedule increasingly unsatisfactory for classifying new inventions, leading to the accellerated update schedule adopted for the 8th and later editions. Some patent offices and databases, but not all, identify the edition of the IPC used in classifying a patent. For large collections of patent documents, the IPC classification definitions in some technologies have been found to be too broad for efficient searching. A hybrid system of classification was introduced for more specific indexing of some patents; countries that use the hybrid system append indexing codes in the format *NNN*:*NN* to IPC subclass codes for certain technologies. The Japanese Patent Office and European Patent Office have attempted to improve the usefulness of the IPC by creating more detailed classifications for use as search tools. The Japanese Patent Office assigns refined classification codes called F1terms, which consist of IPC codes followed by a comma, a numerical subdivision symbol and an alphabetical file discrimination symbol (8). Because variations in patent classification practices can cause difficulties in searching multinational patent files, the EPO has reclassified all of the patents in its search documentation files, using a modified version of the IPC. The European Patent Classification (ECLA) scheme is based on the hierarchical structure of the IPC, but has additional subdivisions to allow more focused searching by EPO staff (9). In databases that include EPO classifications as search terms, an ECLA subgroup may be added to the IPC symbol, in the form of a letter, optionally a number, optionally a letter, etc, depending on the level of subdivision required for the technology. Monthly updates are made to the classification scheme to adapt the classification to technological development, and classifications are changed in the electronic search file so that all patents can be retrieved using the current classification. Beginning with the 8th ed. of the IPC, a similar system will be initiated internationally. Under the reformed IPC the classification of all patents in the EPO's master database will be updated when their IPC changes, and many commercial databases will update their records on a regular basis.

The most common national patent classification codes encountered outside the public search rooms and websites of national patent offices are U.S. classes, which are indexed in many patent databases that include U.S. patents. They are formatted as a one- to three-digit numerical class code, followed by a slash or hyphen and a subclass code consisting of from one to three numbers, which are occasionally followed by a letter or by a decimal point and additional numerals. These codes are also arranged hierarchically according to the scheme published in the U.S. *Manual of Classification* (10,11). Unlike IPC codes, U.S. patent classification codes do not contain clues to their technological relationships.

Patents are given a single original classification and usually one or more cross-reference classifications. The U.S. patent classification system is under constant revision and, as the purpose of the system was to provide a useful arrangement of the patents on the shelves of search rooms, file copies of the patents were moved to their new places at U.S. PTO search rooms until the paper files were replaced by public computer access. Indexes are revised when new classifications are assigned to existing patents. The printed patent specification is not changed when a patent is reclassified, but the U.S. Patent and Trademark Office records the current classification codes of reclassified patents and makes the data available to users of the U.S. PTO databases and to other database producers. Although there is a concordance relating the U.S. classification system to the IPC, the two classification systems are organized according to different principles, and some technologies do not have comparable classifications (12). The U.S. system designation corresponding to the IPC C07c 45/50 would be 568–451 (Fig. 5).

2. Secondary Sources of Patent Information

Patent systems were conceived as a means for promoting technical progress by encouraging the dissemination of information on technological developments. Information dissemination is therefore essential for the patenting process. Patent offices have traditionally announced the issuance of new patents in bulletins and gazettes. Other organizations, notably scientific and technical societies and for-profit publishers, have produced value-added patent information services. These secondary sources of patent information serve multiple purposes, among which are current awareness alerting, document delivery, and retrospective searching. Originally, such products appeared as printed publications, but they are now are used in electronic form in on-line databases. During the 1990s there was rapid growth of optical storage of information, especially as Compact Disk-Read Only Memory (CD-ROM) products; by the middle of the the first decade of the twenty-first century access through the Internet had largely supplanted in-house patent libraries on paper and optical media. Patent documentation is a field in considerable ferment, with rapid introduction of new products, services and capabilities. While most of the databases with controlled indexing systems have existed for many years, databases providing access to the bibliographic data and text of patent documents has proliferated and the relative value of such databases depends on their relative breadth of coverage and search engine capabilities rather than retrieval parameters supplied by database producers.

2.1. Printed and Electronic Patent Office Gazettes, Patent Databases, and Patent Registries. The issuance of patents is announced by patent offices in publications typically known as gazettes and bulletins, which are published most commonly at the time of the patent's publication, but there are exceptions. Advance information is published in a patent gazette by some countries prior to the publication of patent documents, typically as a notification of filing details. However, some patent gazettes do not appear until well after the effective publication date of the patents they announce.

The amount of information included in patent gazettes varies. Typically, they include bibliographic details on published patent applications and granted patents, including patent number, title, inventor, patentee, patent classification, application number and date, and priority application details if relevant. Some gazettes also provide the front page abstract of the patent and a representative drawing. Examples are the *PCT Gazette*, the *Bulletin Officiel de la Propriété Industrielle* of France, and the *Patent Journal*, or *Patentjoernaal*, of South Africa. The *Official Gazette* of the U.S. Patent and Trademark Office includes one or more representative claims. In addition to announcements of new patents and applications, the various gazettes typically include listings of patents that have been rejected, challenged, or disclaimed, patents that have been allowed to lapse, and in some instances even listings of new applications that have been made but that will not be published for some time, if ever. Gazettes often include indexes to the information they contain; the amount of indexing available varies from country to country.

Once available only on paper, patent gazettes of many countries are now published on the Internet, and many patent offices have stopped publishing printed gazettes entirely. The Internet allows the public to access information about newly published patents and applications without mailing delays, and it allows many patent offices to supplement announcements of new patent publications with retrospective databases and continuing updates of the status of those publications. A growing number of countries provide Internet access to bibliographic records, full text, and images of patent publications. Hyperlinks to the databases can be found on a number of patent information websites including those of the Patent Information Users Group and the IPMenu.com site provided by the Australian intellectual property law firm Phillips Ormonde Fitzpatrick. The Intellectual Property Office of Singapore operates a portal to six patent office databases, and the esp@cenet portal, discussed below, provides access to the databases produced by the member countries of the EPO. The status of published and pending patent applications is available from fewer national patent offices as of this writing. The U.S. PTO Patent Application Information Retrieval portal, PAIR, contains records of published applications, granted patents, and unpublished parent applications referenced therein. Accessible without cost or restriction, Public PAIR can be searched by application, publication, and granted patent number only. PAIR records show the progress of the patent application through the patenting process, maintenance fee payments, and patentee contact addresses, and are updated until the patent lapses or expires. Links are provided to the text and image of the published documents. For recent patents, there is also a viewable record of the patent application file, with images of correspondence between the applicant and the office. Private PAIR contains the records of pending applications, with access to each application file limited to its applicants or their representatives. The EPR, EPOLINE, has records of European patents and PCT applications designating the EPO from the date of publication through lapse or the transfer of a granted patent to the national patent offices. Like the U.S. PAIR system, EPOLINE has ongoing status updates, links to patent images, and images of file documents. Unlike PAIR, EPOLINE is searchable with simple bibliographic data such as patentee name and filing date.

2.2. Information from Other Sources. Some of the abstracting and indexing services produced by scientific and technical societies have traditionally included patent information, especially in the field of chemistry. For example, *Chemical Abstracts* (CA), produced by the American Chemical Society since 1907, has always covered patents, as did the discontinued *Chemisches Zentralblatt* and *British Chemical Abstracts*. On the other hand, some notable information services have not included patent coverage. One example, despite the fact that many patents are based on some aspects of engineering, is *Engineering Index/Compendex*. *RAPRA Abstracts*, focusing on rubber and plastics, covered patents briefly from 1978–1980, dropped them for more than a decade, and finally resumed limited coverage in 1994. However, even where patents are covered, the focus may not be ideal for those concerned with the legal aspects of patents. Thus, CA documents the new chemistry disclosed in patents, but shies away from the legal aspects of patents. For these and other reasons, others have stepped in to develop a variety of patent information services, eg, Derwent Information Ltd., which is now a part of Thomson Scientific.

Derwent had its start in the 1950s, when it began publishing abstracts of patents from selected countries: first the United Kingdom, then Belgium, followed by others that included Japan, The Netherlands, and the former USSR. Derwent's country abstract booklets were followed by collections of abstracts covering multiple countries in selected technical areas. These products served an alerting purpose, but had no capability for retrospective searching. During the 1960s, Derwent began a series of more complex information services, providing both alerting and retrospective retrieval capabilities in the fields of pharmaceuticals, agricultural chemicals, and polymers. By 1970, Derwent's coverage was extended to all aspects of chemistry from 12 countries; by 1974, it had begun coverage of nonchemical patents. Over the course of the ensuing years, the Derwent organization has broadened its country coverage and improved its capabilities for information retrieval in many ways. It continues to work on new and improved products and systems, and is the single most important organization involved in patent documentation.

Other organizations have assumed important positions in the field of patent documentation. The IFI CLAIMS Patent Services (formerly Information for Industry) began in 1955 to index U.S. chemical patents by the Uniterm Index system. Uniterm indexing was eventually extended back to 1950. The acquisition in 1971 of Du Pont's in-house indexing system and staff resulted in a more powerful system, the Comprehensive Data Base (CDB), which covers U.S. chemical patents and published patent applications from mid-1964 to date with deeper indexing of chemical structures and linkages between concepts.

Another important resource for bibliographic information on patents is the European Patent Information and Documentation Systems (EPIDOS). The EPIDOS began in 1973 as the International Patent Documentation Center (INPADOC), the joint creation of the Austrian government and WIPO. Subsequently acquired by the EPO, EPIDOS continues to produce the INPADOC database, which at present covers patents from about 75 authorities, and is the most complete source of bibliographic information on patents, including patent family and legal status information. The patents comprising the EPO's search documentation files are available as the worldwide database within the esp@cenet®, a

free service offering limited searching and extensive patent document delivery over the Internet.

There are other organizations providing patent information. L'Institut National de la Propriété Industrielle (INPI), the French Patent Office, is the producer of several important databases and, together with Derwent and the Questel-Orbit databank, has supported the development of the Markush DARC system used in the Merged Markush Service (MMS) to index chemical structures associated with Thomson Derwent's World Patents Index and INPIs PHARMSEARCH databases. The Elsevier Engineering Index EnCompass service, originally American Petroleum Institute's (API) Central Abstracting and Information Service, has since 1964 produced APIPAT, now renamed Ei EnCompassPat, a database covering patents on petroleum refining, petrochemicals, and related technology.

Other specialized patent information products, as well as general information products that include patent and other information, are produced by a variety of organizations, most notably in the area of pharmaceuticals. Unlike most other products, new drugs must undergo extensive regulatory evaluation before they are marketed. Patent terms and terms of market exclusivity for approved drugs are of paramount importance to manufacturers and their potential competitors. Lists of patents relied on for the protection of drugs marketed in the United States are published by the U.S. Food and Drug Administration (FDA) in *Approved New Drugs*, also known as the Orange Book, and the FOI Services Inc. in the series *Drugs Under Patent*. Patents on new veterinary drugs are listed in a corresponding Green Book published by the FDA. On-line searches for patents covering drugs may be performed in the IMSWorld Drug Patents international database. This database contains patent records from many countries on about 1000 marketed drugs and, like the printed lists, is searched by the name of the product. Although these product-specific sources are rather limited in their scope, they provide precisely the patent information most sought by pharmaceutical company executives. More detailed information about pharmaceutical patents is available from Thomson Scientific, in the IDdb3 database of drugs in development, the Current Patents Gazette, and DOLPHIN, "the Database of all Pharmaceutical Inventions". This database uses bibliographic, patent family and status data from the INPADOC database and integrates abstracts, commentary from the Current Patents Gazette, and information from the IDdb3 database. All of these sources are integrated into the Thomson Pharma service along with news and business information.

A growing number of national patent information agencies produce databases containing English language abstracts of their patent publications. The earliest of those databases was Patent Abstracts of Japan (PAJ), which began providing short abstracts of published patent applications in 1976. Patent applications in PAJ include an abstract, written in English as a summary of the complete specification and claims for patents applied for by Japanese nationals in selected chemical, mechanical, electrical and physics technologies with bibliographic information for other patents. The database was originally produced by the Japan Patent Information Organization (JAPIO) and was introduced on online search services as the JAPIO database. *Patent Abstracts of China* is a similar service available covering all patent applications published by the

People's Republic of China since its patent law was enacted in 1985. English language abstracts are produced by the Patent Documentation Service Centre of the Chinese Patent Office for all patents issued to Chinese inventors with bibliographic information for applications filed by non-Chinese applicants. Korean Patent Abstracts, produced by the Korean Institute of Patent Information, has English language abstracts of Korean granted patents from 1979 to 2001 and published patent applications from 2000 onward. Russian Patent Abstracts, produced by the Russian Agency for Patents and Trademarks, has English language abstracts of Russian patents from 1994 to the present.

In addition to value-added databases produced by abstracting and indexing services, there are many databases based on the bibliographic data and text of patents, Originally based on the bibliographic information and abstracts from the first page of patent documents, these databases are increasingly based on the full text of patent documents. A feature of the Trilateral Agreement among the U. S. Patent and Trademark Office, the Japanese Patent Office, and the EPO is an understanding that patent information should be distributed widely and inexpensively. As a result, more patent databases were added to commercial search services, direct acccess to patent data over the Internet was enhanced, and a number of additional services based largely on full text databases were created. Differences among the many databases based on United States and European patents, PCT applications, and Patent Abstracts of Japan lie search software and system features.

3. Advances in Patent Documentation

The last half of the twentieth century saw a strengthening of patent coverage by some traditional abstracting and indexing services whose patent coverage extends back for many years, as well as the establishment of an increasing number of specialty services for the documentation and manipulation of patent information. Advances involved traditional printed products, now largely distributed in electronic formats, as well as various newer electronic forms of information. Computerized databases have become the the dominant resource for users of patent information, and new and modified information tools for retrieval and analysis of patent data continue to appear and develop. Principal patent databases available through on-line databanks and principal Internet based patent information services are listed in Table 2. Single country patent databases, now available from all major online search services and many patent offices, are omitted.

Country coverage of patent databases varies from one database to another; some databases provide complete information about patents published by a single patent-issuing authority, others attempt to catalog the world's entire patent output. Multinational patent databases have historically provided good coverage of heavily industrialized countries and lesser coverage of less industrialized countries. The breadth of coverage and depth of indexing for these databases have depended on the availability of original source materials such as published patent applications, published patent office gazettes, and computerized records provided by the patent offices. Coverage has also been limited by the expense

of translating information from patents in subject areas for which the market is small. Country coverage has tended to expand and contract in response to consumer demand and as the availability of documentation from patent offices changes for political and procedural reasons. A summary of the countries covered by major multicountry patent databases and of the country codes identifying the countries is given in Table 3. Coverage for individual countries may vary in thoroughness with respect to subject matter, type of document, and time span. Users' manuals from such database producers as Derwent and EPIDOS include detailed lists of country coverage ranges. The manuals are available on the Thomson Derwent and European Patent Office websites. Listings of historical patent coverage in Chemical Abstracts are available on the Chemical Abstracts Service website, and most Intranet based databanks also have summaries of country coverage.

3.1. Thomson Scientific. Thomson Scientific integrated the databases produced by Derwent Information Ltd., into a collection of scientific and technical information sources that also includes the ISI databases, and acquired additional patent information resources during the first years of the twenty-first century. As a result, some established patent databases have changed radically or disappeared entirely

Thomson Derwent provides a wide spectrum of information products and services, many of them relating to patents. Derwent also produces important databases (qv) of nonpatent information from the pharmaceutical and agricultural chemical literature. These products and services encompass alerting tools for current awareness, systems for retrospective search and retrieval, and means for document delivery and archiving.

Derwent began as a publisher of simple abstract booklets covering first individual countries, then multiple countries in specific areas of technology, and gradually moved into covering the full range of chemistry (in depth that differs from one subfield to another) and to nonchemical patents. The basic framework of the *Chemical Patents Index* (CPI) was established in 1970, and the overall *World Patents Index* (WPI), encompassing CPI as well as nonchemical patents, was established in 1974. The CPI is divided into 12 sections by technology, as shown in Table 4. The organization of the CPI is essentially unchanged in 2005, although it has undergone many refinements and its country coverage has been broadened considerably. About 41 patenting authorities are covered in 2005, including European and PCT patent publications, and multiple stages of publication are covered for a number of countries. Limitations were placed on the coverage of Japanese patents during the 1970s in response to the sharp rise in numbers of Japanese documents after Japan had adopted the practice of universal publication of patent applications. Chemical coverage was not complete, electrical patents were covered by title only, and general and mechanical patents were excluded. However, Japanese coverage has gradually broadened since then, and coverage of all Japanese patents began in 1996. Not all countries are covered in equal depth. Thus, title-only coverage is provided for some countries, including Italy and the Czech Republic, and abstracts for some other countries are briefer than the Derwent norm. Treatment of general and mechanical patents in the WPI is similar to that of the 1970s except for the addition of more countries; however, coverage of electrical patents in the *Electrical Patents Index* (EPI)

segment of the database has been upgraded considerably with the establishment of a number of subdivided bulletin groupings and the development of an extensive system of manual codes, which have been upgraded on several occasions.

The Derwent patent database is based on records covering a family of equivalent patents. New patent publications are checked against the existing database to see if they are equivalent to previously published references. This is done by comparing priority application details or, if priority is not claimed, by comparing inventor or patentee and technical content with known references. Those publications determined to be new to the system are considered to be basic patents, and are assigned to one or more sections of the Derwent system according to their technical content. Chemical patents may appear in as many sections of the CPI as needed, although a limit of four sections per patent was applied in the past.

Table 4 shows the size of CPI and of its 12 sections. The annual input has more than doubled over the span of 20 years between 1973 and 1993 and increased by an additional 23.3% between 1993 and 2003, and shifts in activity are evident. Thus Section L accounted for just $>11\%$ of CPI references in 1973, but $>26\%$ in 1993; Section D grew from 10 to 16% in 1993 and 30.2% in 2003; and Section G from 6.6 to 11.5% in 1993, but only 8.3% in 2003. Section B, devoted to pharmaceutical patents, rose from $\sim 10\%$ of the CPI in 1993 to $>25\%$ in 2003. The dramatic increases in Sections B and D can probably be attributed to the emergence of the biotechnology industry and related pharmaceutical therapies. The largest relative declines occurred in Sections E and F, while Section A, the largest of all chemical sections, remained relatively constant at about one-third of the CPI. The sectional breakdown of CPI has been important for the marketing of the service. The cost of purchasing coverage for all technologies is substantial, and few organizations have interests sufficiently broad to buy the complete service. By packaging the product in segments, Derwent was able to build up a worldwide clientele. Exclusive access to subscribed sections of WPI has been eroded substantially as on-line access by nonsubscribers has been broadened over the years. As of 1995, Derwent was engaged in reconsideration of its pricing and marketing practices, and it appeared highly likely that shifts away from prices based on section groupings would take place, particularly in view of the fact that pricing of the CPI sections no longer reflected their relative sizes, but as of 2005 a new pricing system had not emerged.

Relatively brief alerting abstracts are written for new basic references. A feature of the Derwent system is the preparation of expanded titles that aim to capture the heart of the invention. The first appearance of the alerting abstracts is in the on-line WPI database, and can be as early as 2 weeks after patent publication for some principal patent offices (United States, EPO, PCT, Germany, United Kingdom); abstracts for other countries, eg, China, take several months to appear. The alerting abstracts are published in several printed products, including alerting booklets and electronic counterparts. Although equivalent patents can be added quickly to the records of existing patent families, the turnaround time from receipt of patent specifications to completion of document analysis varies considerably from patent to patent. The average lag for basic patents from the principal patent offices had fallen to ~ 4 weeks by 2005.

A new database, World Patents Index First View (WPIFV) has been created to allow tracking of incompletely abstracted patents. Patents are added to WPIFV with standardized number and date formats and patentee codes and first page data from the original document. The records are removed when the patents are added to the WPI database.

In addition to the alerting abstracts, documentation abstracts, formerly called basic abstracts, are produced for all basic chemical patents, except for those from the handful of title-only countries. These documentation abstracts frequently provide a substantial amount of technical details beyond those included in the alerting abstract, although for many Japanese references and others from countries such as Brazil, Hungary, Romania, South Korea, and Russia, the alerting abstract and documentation abstract may be identical. Documentation abstracts are produced in a format that highlights key features, including claimed matter, uses, and detailed examples; an example is shown in Fig. 6, which covers the same invention shown in Fig. 2. These abstracts include coding by Derwent's manual code system, which uses a vocabulary of several thousands of keyword-like codes to identify the key aspects of the patent. Documentation abstracts are published in documentation journals covering each CPI section, as well as in profile booklets covering selected segments of polymer and other technology. The text fields added to the alerting abstract are available on-line to subscribers in the extension abstract field. CD-ROM collections of documentation abstracts are produced for archival purposes. Access to documentation abstracts is limited to CPI subscribers; nonsubscribers are limited to the less informative alerting abstracts online and cannot access the extended abstracts online or purchase CD-ROM or printed documentation abstracts.

The manual code system was originally created to enable the searching of classified sets of documentation abstracts grouped by manual code. The system was analogous to the traditional method of searching classified sets of full patent specifications, and has proved to be an effective search method over the years; however, most organizations that once relied on manual code searches have stopped maintaining the card sets because of the cost and space involved in acquiring and filing the cards. Assigned to all chemical and electrical patents except some title-only references, manual codes are included with the alerting abstracts and bibliographic data when references are added to the WPI database, and are a valuable search parameter in on-line searching.

An important advance in the on-line WPI database is the inclusion of representative drawings from patents. These drawings are particularly important in conveying the meaning of mechanical, engineering, and electrical inventions, but are also important in elucidating the chemical structures involved in chemical inventions. The drawings are available online for chemical patents from 1992 onward and nonchemical patents since 1988.

The manual code system covers the entire CPI. In addition, specialized deep coding and indexing systems exist for Sections A, B, C, and E. Section A features the Plasdoc code, introduced at the start of polymer coverage in 1966, and enhanced on a number of occasions since then (13,14). During 1993, Derwent introduced an ambitious new polymer indexing system, that has a greatly expanded indexing vocabulary and the capability to pinpoint contextual relationships among substances and concepts contained in a given patent via a three-

tiered linking system (15,16). Because the linked coding system is fundamentally different from the earlier indexing systems, searching the entire database requires complex time ranging. In Sections B, C, and E, a chemical fragmentation code has existed since the start of pharmaceutical coverage in 1963, and it too has been enhanced several times over the years (17). In 1987, Derwent introduced a topological structure coding and retrieval system aimed at coping more accurately with Markush or generic structures in patents. The topological indexing is applied to both specific compounds and Markush structures and forms a part of the Merged Markush Service on Questel-Orbit. Topological indexing of specific compounds is available as the Derwent Chemical Resource and is searchable within the Derwent World Patents Index database on STN. To supplement their long-standing series of abstract bulletins, Derwent has produced a number of repackaged sets of abstracts. Some of these are targeted at specific industries, such as the automobile industry. Others are created to match the interests of individual companies or groups of companies. Formats of these specialized bulletins can vary considerably. They may include alerting abstracts, documentation abstracts, or even abstracts from other sources; they may be produced as paper copy or in electronic form. This particular Derwent activity is likely to continue in a state of flux as new opportunities are identified and pursued. *Patents Preview* and *World Drug Alerts*, the latter encompasses journal literature and conference proceedings as well as patent information, represent additional Derwent approaches to satisfying the needs for rapid information dissemination in the pharmaceutical industry.

Derwent's comprehensive abstracts can serve as a pointer to patents of potential interest, and in some cases may provide sufficient information to judge the relevance of a patent, but there is no true substitute for the examination of complete patent specifications, particularly when legal decisions must be made. An aspect of the Derwent service since its early days has been the provision of microfilm and, in selected instances, paper copies of patents covered by the system. In most instances, the basic patent and key English-language equivalents were included. English abridgments of Japanese patents were included until the early 1970s, when the volume of Japanese documents produced under the then-new Japanese patent law made this impractical. Complete patent specifications have been provided in groupings by Derwent section. Paper and microfilm patent archives have largely been replaced by electronic document delivery on demand. The Thomson Patent Store is a document delivery service for worldwide patent copies, and Thomson Derwent operates a translation service for non-English patents. Document delivery services are also a major component of Thomson's three Internet-based patent search services, Delphion, Aureka, and the MicroPatent PatentWeb.

Derwent has developed, either on its own or through contractors, a number of computer aids to information processing. The TOPFRAG series of programs, available as MARKUSH TOPFRAG, aids users in searching chemical structure information including the generation of search strategies for the Merged Markush Service and for fragmentation code searching of the CPI. The program MARKUSH TOPFRAG has also been embedded in STN Express software as an aid in searching structures on Derwent files in the STN system. The

PILOT program is for assisting searchers with strategies using the polymer indexing systems.

The Derwent Patents Citation Index (DPCI) offers an alternative to searching patent or abstract text and subject matter searching. The DPCI is an on-line database of patent citations that includes examiners' citations to prior art from patent specifications. When given a known invention of interest, as represented by a WPI patent family, the database can identify any patent against which it or a family member has been cited, as well as earlier patents cited by any member of that family. When the database was originally released, citations from the patent disclosure were also included, and those cited references are also in the database. Limited citation searching capability is available in a number of databases, but Derwent's file greatly increases the ability to carry out citation searches by consolidating citations to and from any of a family of patents claiming the same invention.

Thomson Derwent also produces a biotechnology database, Biotechnology Abstracts that covers both patents and journal literature, and a biosequence database GENESEQ, that indexes sequence structures of proteins and nucleic acids disclosed specifically or generically in patents. This database is available for in-house searching searchable with special sequence software on the IntelliGenetics system, and as the DGENE database on STN. The Journal of Synthetic Methods contains organic chemical reactions from patents as well as the journal literature.

3.2. Chemical Abstracts Service. The Chemical Abstracts Service (CAS), a division of the American Chemical Society, has produced *Chemical Abstracts* (CA) since 1907. Since the demise of *Chemisches Zentralblatt* and *British Chemical Abstracts*, CA has been the preeminent medium for documenting new publications in the field of chemistry, biochemistry and chemical engineering. The CA documents chemical publications of all types. It is not a patent database per se, but its patent component is larger than most databases devoted entirely to patents. Thus, for example, the number of patent references in CA for the years 1991–1993 ranged from 95,500–99,400/year, but had increased to 164,400 by 2003.

Derwent products have always been targeted at those in and near the legal profession, and the company itself has always attempted to emphasize in its treatment the fact that patents are legal documents. Chemical Abstracts, on the other hand, has the mission of documenting chemistry for chemists and chemical engineers. Therefore, CA abstracts of patents have emphasized what was actually done in examples that provide hard data, and have avoided discussing the purpose or scope of a patent or prophetic paper examples. The CA abstract for the same patent whose Derwent abstract appeared as Fig. 6 is shown in Fig. 7. Until $\sim$1980, CAs abstracts and indexes concentrated only on examples, not claims; however, from about 1980 onward, CA has been indexing substances covered explicitly in claims even if these are not described in examples with hard data. Another significant change since 1988 has been the structural indexing of Markush structures from patents in the MARPAT database.

Just as Derwent has broadened its country coverage, so has CAS. During the 1960s, CAS policy was to cover all chemical patents for only a handful of countries; for other countries, only patents to nationals were covered. The

group of countries changed from time to time, and at times only the United States, the United Kingdom, and Germany were fully covered. French coverage was theoretically also complete, but in practice less than that. That was the era before the widespread introduction of the practice of publishing all patent applications. The United States, the United Kingdom, and Germany published only those applications determined to be patentable, and one consequence of the CAS coverage policy was that many inventions that appeared only in countries where coverage was limited to nationals did not get abstracted or indexed. Beginning around 1968, the number of countries afforded full coverage for chemical patents has increased, so that CAs coverage of chemical patents is far more complete in the 2000s. Coverage now extends to 48 patent issuing authorities. Only twelve of these countries are limited to nationals-only coverage.

Chemical Abstracts began its existance in 1907 as a printed publication and its bibliographic records and indexing became available as an online database, CASearch, during the early years of online search services. CASearch has records for literature and basic patents from 1967 to the present and was made available on most major online search services. A more complete database, the CA File became available with the creation of the STN International search service in 1984, and additional data including abstract text, representative chemical structure drawings and links to the CAS Registry filewere made available. As an added enhancement, CAS created the CAPlus database, which contains bibliographic information and author abstracts for publications such as ACS meeting presentations that are not included in the printed CA and incompletely indexed records for references for which indexing is not yet complete. Unlike the printed CA and CASearch, patent family information is available in the STN CA databases, with family information extracted from the INPADOC database and added to CA records retrospectively. Abstracts from the earliest years of Chemical Abstracts have been scanned and added to the database and cited references from abstracted documents have been added to the records. In addition, CAS is adding bibliographic records for journal articles published prior to the inception of Chemical Abstracts and cited in later publications. Before the enhancement of the STN files, there was an important difference between the CAS and the Derwent treatment of members of a patent family. Derwent abstracts the first member of a family that it sees, then adds bibliographic data for all equivalent patents to the record, so that the record for a given invention in the WPI database can be accessed by bibliographic information on any member of the family, including patentee and inventor names that differ from those on the basic patent, and the family information is available in all versions of the WPI database. Although CAS also abstracts the first member it sees of a family, it enters subsequent equivalents into only the printed CA patent index and the STN CA and CAPlus files, but not the CASearch database licensed to other search services. Thus a searcher looking for a given patent number in CASearch will find it only if it was the member of a family that happened to be covered first. On services other than STN, patent family information for CASearch records can only be achieved via the cross-file techniques of various on-line systems.

3.3. EPIDOS (Formerly INPADOC). The International Patent Documentation Center (INPADOC) was created as a result of agreements reached in 1972 between WIPO and the government of Austria. It reflected the desire

of many in the intellectual property community to have an authoritative and noncommercial repository and dissemination center for bibliographic information on patents. The INPADOC operation has now become part of the division of the European Patent Office known as the European Patent Information and Documentation Systems (EPIDOS). Information is obtained directly from national and international patent offices, which as of 2005 number around 75 and include more countries than any other patent information service. Title, inventor, patentee, classification, as well as priority, application, and publication details are all included for different stages of publication, including in some cases unexamined, examined, and granted patents, and even unpublished applications for some countries. INPADOC ties together members of extended convention patent families, and in some instances identifies intellectual families as well. Information about legal status subsequent to publication is also collected for a growing number of authorities, although the time range and completeness of coverage varies considerably among the countries covered. EPIDOS originally issued printed and microfiche compilations of INPADOC data; its database can be searched on several on-line host systems and has been integrated into various other patent search services. In general, EPIDOS provides the most complete patent family information of any service, although Derwent tends to include more information on intellectual (nonconvention) families.

Independently of the INPADOC organization, the EPO created a database of patent families for the use of its patent examiners. The DOCd.b file compiled patent families representing the patents in the PCT Minimum Documentation files. documenting the availailability of a family member using Patent documents going back to 1920 and beyond have been reclassified according to the IPC system and the EPOs modified ECLA classification systems, and a representative member of each patent family was digitized. The digitized patents and family information are available to the public as the Worldwide database on the internet based esp@cenet service. EPIDOS also carries out individual searches and provides patent copies as well as information on Japanese patents, including English-language abstracts and searches of the Japanese-language PATOLIS database. Another service from EPIDOS is the series of ESPACE CD-ROM products, providing document delivery of full patent specifications from the EPO, PCT, and a lengthening list of individual countries. Approximately 1000 full specifications can be contained on an individual CD-ROM for printed documents such as European or U. S. patents. PCT applications, typed with wide spacing between lines, require more storage space. Other ESPACE CD-ROMs provide in searchable form the front pages of European and PCT patents, the *EPO Bulletin,* legal decisions, and other information of interest with regard to intellectual property. The CD-ROM products documenting patents less developed contries are also produced.

Esp@cenet has become one of the most useful patent search services. It has a rather simple search engine, with only patent titles, abstracts, and bibliographic data searchable, but it provides patent document images and limited family information at no charge. Links to esp@cenet are widely used by other search services to obtain patent documents, and a number of vendors provide software for collating individual pages into full PDF document images. The EPO interface to esp@cenet is searchable in English, French or German and

contains, in addition to the Worldwide collection, files of European and PCT applications and Patent Abstracts of Japan. Esp@cenet also has search interfaces from each of the EPO member countries, each in the national language or languages. Each national esp@cenet site provides a database of at least 2 years of its national patent documents in addition to the Worldwide collection. In addition to a national Spanish exp@cenet database, the Oficina Española de Patentes y Marcas provides the Latipat database of patents from Latin America, with Spanish and Portugese interfaces to patent publications from 19 Latin American countries. As of 2005 information from most of these countries is incomplete, but better coverage is promised in the future.

3.4. IFI CLAIMS Patent Service. The IFI CLAIMSs predecessor company, Information for Industry (IFI), began indexing U.S. patents by its Uniterm system in 1955. Coverage was eventually extended back to 1950. In 1972, the Uniterm system was complemented by a more powerful retrieval system based on a merger with an indexing system developed by Du Pont and acquired by IFI (18,19). The latter system, called the Comprehensive Data Base, is available only to subscriber organizations. With the advent of on-line databases, these chemical indexing systems were augmented by bibliographic information, including bibliographic data for nonchemical patents going back to 1963.

The IFI has never been a patent abstracting organization, although it does provide subscribers with sets of bibliographic, abstract, and claim information selected from U.S. PTO tapes to match the subscribers' interest. Besides its on-line databases, IFI produces magnetic tape versions of the databases which some users choose to run in-house. Other IFI patent products include the *Patent Portfolio Service* for manaagement of corporate patent documentation, and the *Patent Intelligence and Technology Report*, an annual listing of patents to all organizations receiving at least 10 U.S. patents during the preceding year. These reports show the total number of patents, and a breakdown by U.S. patent class. One shortcoming of these reports, as well as of patent count lists issued by the U.S. PTO itself, is the fact that they do not aggregate the patents for those industrial organizations that choose to have their patents assigned to multiple subsidiaries or divisions. Some corporations receive patents through 20 or more subdivisions, whereas others use a single patenting entity. Comparisons made on this basis can thus be quite distorted, but IFI's standardization of company names helps to reduce distortions based on spelling, punctuation and abbreviations, resulting in statistics that differ significantly from those based on raw patentee name data.

3.5. L'Institut National de la Propriété Industrielle. The French Patent Office (INPI), is a principal provider of patent and trademark databases, all of them accessible on Questel-Orbit. FPAT, EPAT, and PCTPAT have full bibliographic data, abstracts, and claim text for French, European, and PCT patent documents, including information about changes in the status of the applications after their original publication. The text of the European Patent Classification system (ECLA) is searchable in the ECLATX file, which is updated monthly with changes made to accommodate changes in technology reflected in patent applications. PHARMSEARCH is a structure-searchable file of pharmaceutical patents from France, the EPO, the PCT, the United States, the United Kingdom, and Germany (20). The database is searchable through the Markush DARC

system and has extensive indexing of pharmaceutical concepts, abstracts, and displayable images. The original database was INPIs pharmaceutical structure search file, and patents in the file were indexed retrospectively; French special medical patents published between 1961 and 1973 have been indexed with Markush DARC, thus adding structure-search capability to a small segment of the early patent literature. In 2000, INPI discontinued subject-based indexing of patents and now concentrates on retrospective chemical structure indexing of patents from the Derwent World Patents Index for the Merged Markush Service.

3.6. Elsevier Engineering Index. The American Petroleum Institute's Central Abstracting and Information Services were a unique example of the creation and molding of a series of information resources, including a patent database, by a group of companies with similar interests (21). During the 1950s, it was common in the petroleum and petrochemical industry for individual companies to have in-house abstracting and indexing groups. As costs of such operations increased, these companies turned to the American Petroleum Institute (API) to provide a vehicle for centralized production of information services. Literature coverage and a specialized bulletin on Soviet literature began in the 1950s, and a printed patent abstract bulletin began in 1961.

During the early 1960s, cooperative efforts under the API resulted in the development of a system for indexing these bulletins for searching by computer. The heart of the indexing system was a thesaurus developed by a study of a year's worth of published literature in the fields of interest. The resulting APIPAT and APILIT databases were launched in 1964, and were among the first databases to go on-line in the mid-1970s (22). The databases continue to be available as EI EnCompass.

In 1972, the API reached an agreement with Derwent to use repackaged Derwent alerting abstracts for its printed patent bulletins and to do its patent indexing from Derwent documentation abstracts. This enabled API to discontinue patent abstracting for the most part, and the documentation abstracts provided richer material for indexing than did the relatively brief API abstracts. Cross-referencing from the APIPAT database to WPI enabled on-line searchers to move from hits in the APIPAT database to the corresponding WPI references, including their complete patent families. Ultimately, the APIPAT and WPI databases were merged on the Questel-Orbit system, which enables a searcher to combine EI EnCompassPat and Derwent retrieval parameters in a search (23). EI EnCompassPat and WPI remain separate databases on the DIALOG and STN systems.

Perhaps the most notable aspect of the history of the API operation is that it was been shaped at every step by those who use the system. Created by information specialists within the petroleum industry, it has been governed by a technical information committee made up of company representatives, and guided by a series of industry task forces, which have modified as needed the indexing thesaurus, subject selection guidelines, and selection rules for countries in patent coverage, journals, and other sources in the nonpatent literature. The task forces continue to operate for the EI EnCompassPat database after its transfer from API to Engeneering Index.

4. Types of Patent Information Searches

There are many different reasons to search for information about or related to patents. The methods, sources, and techniques vary widely, depending on the purpose and the complexity of the individual situation. Searches can vary from free to well over $5000. It is essential when searching for patent information to gear the strategy to the task at hand. An ill-conceived computer search strategy can produce mountains of output that can require a huge outlay of time and or money to analyze. An inadequate search strategy can lead to even greater costs if the result is patent infringement. Anyone performing patent searches must have a sound understanding of the costs and benefits that may be involved.

4.1. Novelty Searching. At the heart of the patenting process is the novelty or patentability search (24). A novelty search should be carried out by an inventor or an inventor's representative before a patent application is drafted in order to help ascertain whether the invention is indeed patentable and, if so, what its limits might be. A novelty search is carried out by a patent office examiner to make a decision on the patentability of an invention. Although there may be in some instances a temptation for an inventor to omit or ignore the novelty search and rely on the examiner's work, the cost of making a patent application is considerable. Further, the failure to be aware of relevant prior art when filing an original application can place limitations on the applicant's ability to reshape the application by amendments. A novelty search is normally focused sharply on the specific details of the invention, but broader searches that provide a context for the invention in relation to the state of the art can be justified when an invention gives promise of having wide application and high value.

A valid patent covering any claimed invention can be obtained only if the wording of the claims defines an invention that has never been used or described before the filing of the patent application, and that is not an obvious variation of something that has been described in the prior art. However, the standard for prior art references that may be brought to bear against a patent differs from country to country. Thus, for most countries, a standard of absolute novelty applies, ie, references can come from anywhere in the world and can have been published at any time prior to the priority filing date of the patent application. For other countries, which are far fewer than in the past, only references that reside physically within the country are considered. Exceptions are made in some countries for public disclosures by the inventors during a short grace period prior to the filing of the application claiming the invention, or for public use of the invention for experimental purposes. In the United States, which grants patents to the first inventor rather than the first to file a patent application, a description of the invention published less than a year before the application was filed may be discounted as prior art if the applicant can prove that the invention was made prior to the publication date. Standards for judging whether an invention is unpatentably obvious also differ from country to country. Most countries will not grant a patent covering a claimed invention that is similar enough to one described in a previous publication so that a person working in the field would consider the differences trivial. Thus, a publication showing a process using ethanol as a solvent would prevent in most countries the grant

of a patent on a similar process that uses methanol. Most countries consider an invention to be unpatentable when information from two separate publications can be combined to match the claimed invention. Standards for judging whether an invention is unobvious or has an inventive step over the prior art are based on precedents set in published decisions of the patent office or national courts.

Traditionally, novelty searches have been performed by leafing through stacks of patents in those divisions of a classification that seem best to categorize the invention. This sort of searching is facilitated greatly by the front page of the modern patent specification, which can often show the searcher at a glance whether or not the patent is likely to be relevant. It requires that the searcher have access to appropriate sets of classified patents and that the integrity of the collection of patents is maintained: in the twenty-first century such collections are universally available in electronic form. Further, it excludes from consideration those patents that might be relevant to the search but which, for hierarchical or other reasons, are classified elsewhere. For example, two patents may be closely related. One may be classified under the product and not cross-referenced to the process, the other classified under the process but not cross-referenced under the product. As a result, the earlier patent would not have been considered when the later one was granted. This is because an examiner may cross-classify a patent, which has become more and more the common practice, but legally the examiner is required only to classify a patent in the most relevant class. Thus, searching based solely on classification risks the omission of useful references.

Searching of one or more commercial on-line databases is a technique increasingly used in novelty studies. The use of such databases enables the searcher to combine indexing parameters, including national and international classifications; natural language words in the full text of patents, in their claims, or in abstracts supplied by inventor and by professional documentation services; and indexing systems of various sorts. Because the various patent databases have strengths and weaknesses that complement each other, the use of multiple databases is thus prudent, and is facilitated by multifile and cross-file techniques provided by the various on-line hosts.

On-line searches carried out in this way can provide impressive recall of potentially relevant documents, but depending on the search strategy used the results can be quite dilute; they can include many references that are not relevant. The evaluation of the results of on-line searches can be more difficult than the evaluation of a hand search through classified sets of printed patent specifications. The searcher in general obtains a computer printout or a listing of hits from an Internet search, which may be quite lengthy and many of whose listings may provide insufficient information to determine relevance. Output from a simple bibliographic database such as INPADOC provides no subject information beyond a title, often uselessly brief, and patent classes. Abstracts from CA or WPI can be more helpful in determining relevance, but can still leave much open to question. Derwent documentation abstracts and their rich information content are useful as a primary search output, and CA indexing text is often an excellent pointer to chemical information in patent. Documentation abstracts and in-depth indexing terminology, useful as they are, may also prove inadequate for final decisions, which would make it necessary for the searcher to obtain and examine copies of full patent specifications. Original author abstracts

from patent documents are usually even less informative. Patent abstracts must inevitably be treated as tools for screening potentially relevant patents from false retrievals. The increasing availability of full patent text and patent images through the Internet and document delivery services has largely removed the obstacles to evaluating patent search results.

Novelty searches are not necessarily limited to patent information. The anticipation of a purportedly novel idea can occur in journals, books, magazines, etc. Thus, the potential scope for a novelty search is essentially infinite, and one of the challenges to the searcher is to devise an effective strategy whose cost is commensurate with the potential value of the invention.

4.2. Infringement Searching. An individual or organization found to be infringing the patent rights of others is subject to penalties that can be extremely costly. It is essential for anyone contemplating a commercial venture that is technology-dependent to find out first whether or not the proposed venture falls within the area covered by adversely held patents.

Whereas the potential field of search for the novelty search is essentially limitless, there are certain limits that can be placed on an infringement search. References, either new or old, relevant to a novelty search can appear in any medium, patent or nonpatent, anywhere in the world; they may be found in claims or in disclosures anywhere in a patent specification. By contrast, an infringement search can be limited to the content of the claims of patents, and only to the country or countries in which manufacturing, sale, or use of the invention is contemplated. Only patents that are in force or that are potentially in force need to be considered. Patents that have expired, that have been invalidated, or that have lapsed because of failure to pay maintenance fees can be excluded from consideration. A searcher must be alert, however, to patents that are potentially in force. Thus, if the United States is the country of interest, attention should be given to patent cases that have been published in other countries and are likely to be pending in the United States. If the intention is to operate in one or more countries outside of the United States, consideration must be given to published applications that have not passed examination, but that might be granted in the future. Also, a patent on a given substance might exist without composition of matter claims but include process claims that in themselves are not of concern. The searcher must also consider the possibility that a divisional patent with composition claims might yet be issued.

Since the exact language of claims is vital to matters of infringement, the search of full patent specifications remains the most reliable method of infringement searching. The full claims text of all United States, European, German, French, Japanese, and PCT patent documentss is available in many on-line databases covering a time span longer than the life of a patent. Although bibliographic information and abstracts are available for all industrialized countries, there are many countries whose patent claims are not yet searchable on-line. On-line databases are sometimes used for infringement searches by carrying out careful searches of parameters other than the claim language. However, reliance on computerized databases lacking full claims text for infringement searches involves compromises; at the very least the searcher must obtain the full claims text, including any associated drawings or chemical structure diagrams, of all patents of potential interest that are disclosed by the computer search.

4.3. Validity and Opposition Searches. Given the identification of a patent that presents a potential infringement risk, an individual or organization may choose to obtain a validity study in the hope that references can be located which show that an invention was either anticipated or obvious, and that the patent should not have been granted. As was the case with novelty searches, the potential scope for a validity search is broad. It can include both patent and nonpatent literature. In particular, a disclosure in a patent specification not closely tied with that patent's claims can often be useful in invalidating a patent. Since such disclosures are typically not reflected in the classification of the patent, which is tied to the claims, patent classifications alone are not necessarily effective for validity searching. Deep-indexed databases on the other hand can be useful, and full-text patent databases also have great utility in validity searching. The stakes involved in gaining freedom from blocking patents can be substantial; the cost and effort expended in a validity search can be correspondingly large.

Closely akin to validity searching is searching for the purpose of opposition. Long a factor outside the United States, this technique is becoming increasingly important in the United States as companies engage more and more in worldwide operations. In most countries outside the United States, when the examiner has been satisfied that an invention is patentable, the patent specification is published either as an examined application or as a granted patent, and third parties are given a limited period during which to oppose or challenge the patent. Validity and opposition searches have the same requirements as novelty searches, ie, any reference that would render an invention unpatentable under national laws is relevant in an opposition search. Unlike novelty searches, however, opposition searches are performed long enough after the filing of the patent application so that all of the prior art published before the date of filing is made available for searching.

4.4. State-of-the-Art Searches. State-of-the-art searches are typically carried out when research in a newer area is to begin in order to identify what has previously been done, what is known, and where fruitful opportunities might be found. Typically, a state-of-the-art search is broad and general, although tighter and more focused follow-up searches are often carried out once the areas of potential interest are identified. Detailed state-of-the-art searches are not the norm, but it is quite common for organizations to prepare and maintain such detailed studies in subject areas of great importance and significant commercial or research interests. It is also possible to do a very broad search of companies or technologies of interest and use text- and data-mining tools to map the content of the prior art.

4.5. Alerting Searches. Various means are available for keeping up with the latest in patents, and it can be effective to use a combination of these methods. Thus, computer profiles created to represent individual, group, or organization interests can be run against databases as they are updated, and specific searches can also be run against these databases. National and international patenting authorities are increasingly publishing patent documents on the Internet rather than on paper on the date of publication. Internet databases without controlled indexing can be updated on the date the patents and applications are published or within a few days of the official publication date. The Current

Patents Gazette is issued within a week of publication. Other value-added databases tend to be slower. English language abstracts in the Japanese, Chinese and Russian patent abstracts databases lag publication by as much as six months. Derwent WPI data, using value-added access points, is on-line an average of 30 days after patent issuance from principal patent offices; this time is somewhat longer for other countries. The World Patents Index First View is an attempt to close the gap by providing first page data from basic patents that are not yet ready for inclusion in the WPI database; the records are provided in a rolling file with patents moved to the WPI file when indexing is complete. The amount and kind of data provided for the patents is highly variable; for example, machine translations of abstracts are provided for Japanese and Korean patents, and German language documents have German language abstracts. While patentee codes are assigned and numerical data is standardized, the database is useful only for simple current awareness searches. Chemical Abstracts is also variable in its timing; some patents can be covered rapidly, others lag substantially. The CAPlus database on STN helps in providing an advance look at unpolished abstracts and indexing slated for the printed version of *Chemical Abstracts* and the CA Search file licensed to other search services. The use of machine translations makes inclusion of Japanese patents in CAPlus particularly timely. Similarly, the MARPAT Previews file gives early access to Markush structure information from incompletely indexed patents.

Other printed and on-line products are aimed at alerting in targeted areas, especially in pharmaceuticals. Thomson Current Patents produces the Current Patents Gazette with short abstracts of U.S., British, PCT, and European patent applications within seven days of issue for the *Current Patents Gazette*. These abstracts emphasize the pharmacological and pharmaceutical aspects of the patent disclosures rather than the legal content of the patent. Derwent publishes patent abstracts from a larger group of countries in a similar format in *Patents Preview* and *World Drug Alerts*, which feature the novel aspects of the inventions as well as their pharmaceutical utility. These abstracts are different in format from CPI abstracts and are distributed before the corresponding CPI record is created for printed or on-line access. When abstracts are prepared within days or weeks of patent issuance, the speed of alerting services hinges more on delivery routes than on production rates. These services are also available in electronic form for access within subscriber organizations, and updates can now be transmitted electronically, thus avoiding mail delays.

Traditional browsing through patent office gazettes and abstract bulletins still serves a useful purpose in patent alerting. It can be difficult to frame a query for a computer search on all the subject matter that might be of interest to an organization. The human mind can spot unanticipated material and relate it to interests, something presently beyond the power of the computer.

Most new patent cases of interest are published by at least one of the United States, European, or Japanese patent offices, and WIPO (PCT). Japan presents problems for those not able to read Japanese, but the U.S. *Official Gazette* (with representative claims) and *PCT Gazette* (with English-language abstracts) are on the Internet on the date of publication, as is the *European Patent Office Bulletin*, which contains trilingual titles, and electronic copies of the patent documents themselves are available for more detailed review. When a patent application

is known to be of interest, the PAIR and EPOLINE registries can be monitored to see when a pending patent application is scheduled for grant. A highly effective alerting program can be developed from a combination of these methods.

4.6. Family and Equivalent Searches. A wide range of inquiries fall under the category of patent family searches. It may be desirable to find an equivalent to a known patent in a given language, typically but not necessarily English. It may be necessary to find whether an invention is protected in a given country. It may be desirable to estimate a patentee's interest in an invention on the basis of how broadly it has been filed, or to know in detail all the countries in which an invention has been patented, including the legal status of each (25). Or it may be necessary to trace the entirety of a complex extended family, replete with divisionals, continuations, continuations-in-part, and multicountry equivalents at one or more stages. All of these tasks have become relatively simple because of the efforts of Derwent, EPIDOS, and CAS. Derwent's WPI database covers 41 patenting authorities in 2005, and provides information on multiple stages of publication in many of them. It identifies many intellectual patent families, and provides data links that make tracing the web of extended families possible. Family information goes back to 1970 for chemistry and several years earlier than that for pharmaceuticals, agriculturals, and polymers. The Questel-Orbit version merged with Ei EnComPassPat also includes family information from the 1960s relating to petroleum and petrochemicals.

In the INPADOC database, EPIDOS covers even more countries and stages of patent publication, and includes some patent status information for ~ 40 authorities. Coverage for most major countries goes back to 1968 or 1973; starting times for other countries vary. Family searches in the INPADOC database identify extended families. EPIDOS identifies some intellectual families, though fewer than Derwent; however, it is often possible in searching INPADOC to identify nonconvention family relationships. Prior to its acquisition of the INPADOC database, the European Patent Office collected patent family information in the Doc.d.B database, which included older patent family data than INPADOC. The DOC.d.b data is now incorporated into the esp@cenet world database and the Questel-Orbit FamPat/PlusPat databases. These files combine data from INPADOC, DOCd.b and national patent databases on the Questel-Orbit system and are virtually unique among non-Japanese language databases in including information on C-stage Japanese patents, ie, those that successfully weathered the pregrant opposition period and been sealed as patents under pre-1996 patent law. It also contains some information on patent family relationships from the period long before the advent of patent family databases. A similar combination of sources is available on the subscription Minesoft PatBase service.

An often forgotten source for some patent family information is the CLAIMS database. Although its direct coverage is limited to U.S. patents, it includes limited patent family information up to 1979, for Belgium, France, Germany, The Netherlands, and the United Kingdom. Some of this information from the pre-1970 era is found in no other on-line database.

The *Chemical Abstracts* database has an on-line family capability in the STN CAPlus files with family data obtained from INPADOC. Printed patent indexes have been including family information since the 1960s, but the number of countries covered before 1970 was limited. The incorporation of family

information in the online file has erased this shortcoming in CA patent coverage and has turned STN into the search service of choice for patent searchers.

4.7. Citation Searching. In the scholarly literature, authors cite earlier publications that relate to the work being reported, thus a subject relationship exists between the citing and cited literature. This relationship has formed the basis for the *Science Citation Index* and related products, developed by ISI, the Institute for Scientific Information, now a unit of Thomson Scientific. Known as Scisearch in its on-line version, the *Science Citation Index* is an important information retrieval tool. It can be used for straightforward subject searching, in which mode it complements traditional indexed databases and indexes. It is a popular tool for bibliometric studies of various sorts, such as attempts to measure the relative impact of research carried out by different individuals or organizations, or the relative impact of publications in different journals. Science Citation Index is not a patent database, but it does record patents cited in the journal literature.

Citations appear in patents as well as in the journal literature, and it has been proposed that they, too, can provide different types of useful information, such as research trends and estimates of the effectiveness of research organizations. Elaborate techniques for using citation data for such analyses have been developed (26,27). However, citations found in patents differ in a number of ways from literature citations and these differences can strongly color search results (28-30).

At least three types of citations in patents can be identified: inventors' citations found in the patent specification, examiners' citations found on issued U.S. patents, and examiners' citations found on published applications and granted patents from other countries. A patent inventor cites prior art in order to distance the invention from that art, rather than to show a close relationship. Whereas scientific researchers may want to show how closely they have built on what went before, for an inventor that can suggest anticipation or at least obviousness. Thus, citations within a patent typically try to demonstrate the inadequacies of prior inventions and the uniqueness of the patentee's own work. References tied by this type of citation can be useful in developing a picture of the state of the art, but often show sharply differing technologies.

An examiner's citation found on a granted patent should show art that is related to the invention at hand, but which did not anticipate that invention. If the invention had been anticipated, the patent would not have been issued. When a patent is an improvement patent on an earlier invention, examiner's citations typically show the fundamental invention and represent an analogue to the traditional literature citation. In addition, examiners' citations often are used to show general background and the state of the art. Although patent office procedures require the citation only of prior art relevant to the pending claims, there is a tendency among U.S. examiners to cite dozens, even more than 100 earlier references, even though long citation lists dilute the meaning of citations. Many of these questionable references are submitted by the applicant under what is known as the duty of candor, which requires that applicants cite all pertinent prior publications known to them in an Information Disclosure Statement. Furthermore, some examiners appear to have personal favorites that they cite whenever a given subject area comes up, regardless of how closely the technologies

match. Both of these factors diminish the significance of some citation searches of U.S. patents.

A distinct difference between examiners' citations on granted U.S. patents and those on published patent applications is that the latter can indeed represent direct anticipation. Thus they represent a close subject relationship to the document in question. An important factor in the citations on EPO and PCT applications is that they are categorized by the examiner with regard to their relevance: documents of particular relevance in themselves, documents of particular relevance in combination with some other document(s), and documents defining the general state of the art but of no particular relevance in themselves. Clearly not all citations have the same value.

Citation searching of patents offers a perspective different from either traditional class searching or traditional subject index searching. A citation search on known fundamental patents can lead directly to improvement patents, even when those patents are so new that they have not yet been indexed. This technique can be especially effective when working in an unfamiliar area, or one which is difficult to index.

The availability of citation searching tools has increased enormously. A citation database for U.S. patents was first built by Search Check, Inc., which collected all U.S. patent citations back to their first appearance on patent copies in 1947. This database was subsequently made available for on-line searching as the CLAIMS-Citation database. Lower priced citation searching of U.S. patents from 1971 onward has been available since the advent of front page and full text U.S. patents files, which are now available on all major search services and many specialized subscription patent services. Hyperlinking of cited patents on Intranet- based patent databases greatly facilitates review of older and newer patents on a technology. Citations are also available in databases covering European, PCT (WO) and German, databases. Examiners' citations in early EP and WO patent documents are available in the WPI database. In 1995, Derwent introduced the Patents Citation Index (DPCI) covering inventors' and examiners' citations from 16 countries, and enabling searches to be carried out for citing patents as well as for cited patent and nonpatent references. Although the number of countries has been reduced and inventors' citations are no longer indexed, the availability of this tool sharply increases capabilities for citation searching, and provides better exploitation of an intriguing type of patent information. Nearly all patent citation resources are limited in the number and kind of cited references they recognize, most commonly U.S. patents cited in other U.S. patents. The DPCI is unique in indexing the WPI accession number for each cited and citing patent, allowing the searcher to use crossfile searching techniques to create a single listing of cited inventions. Thomson has created a subscription extranet service, the Derwent Innovations Index, with documentation abstracts, hyperlinks to cited and citing patent families and the possibility of linking many non-patent citations to the Web of Knowledge portal to the Science Citation Index.

4.8. Business-Related Searches. Many searches of business-related questions can be answered by searches of patent information. For example, an organization may wish to study the patent assets of competitors in a technical area or to evaluate similarities and differences in approach and strategy between

its own and other organizations. Statistical analyses based on citations and other data may be desired. Searches may also be desired to identify candidates for joint ventures or for acquisitions or divestitures, or to clarify the relationships of corporate segments. Knowledge of the technology behind new product or process announcements by competitors, or the technology being offered for license or purchase by an individual or small organization, may likewise be needed. Searches of patent databases are invaluable in answering all of these types of questions.

5. On-Line Database Searching Methods

5.1. Coordinate Indexing and Boolean Logic. Three methods of indexing have been prominent in the chemical literature. The first, articulated indexing, has been used in printed *Chemical Abstracts* subject indexes from their earliest days. A number of important concepts are identified as permissible index entries, including specific compounds, material types, reactions, and processes. One or more modifying statements follow each basic index entry. Thus, eg,

Hydrocarbons, pyrolysis of, in plasmas

A second type of index, the keyword-in-context (KWIC) index, arose during the early days of computer processing. The same entry would appear in a KWIC index as follows:

PYROLYSIS OF HYDROCARBONS IN PLASMAS

The third type of index is the coordinate index, in which all of the individually indexable concepts of a document are posted to the record for that document. Entries in coordinate indexes may be based on groups of words, single words, or codes. Coordinate indexes became significant during the 1950s and 1960s in such tools as the Uniterm Index to U.S. Chemical Patents, as well as in personal information tools such as optical coincidence cards and edge-notched cards. The aforementioned index entry would produce three indexing terms:

HYDROCARBONS, PYROLYSIS, and PLASMAS

In the earliest days of on-line databases, all three indexing types collapsed into the third. Using older manual tools, it was difficult to coordinate more than two or three concepts, but the computer made that easier. Each concept in a search can be represented by a string of synonyms or alternatives, and searching can be done for two such parameters or more. Thus, Boolean logic expressions can easily be constructed as follows:

(A1 OR A2 OR A3) AND (B1 OR B2 OR B3) AND
(C1 OR C2 OR C3 OR C4) AND...

However, this advance has an important shortcoming: the lack of context. More than one idea is expressed in a document; a patent on oxidation catalysts, for example, could include examples of the oxidation of methanol to formaldehyde and of 2- propanol to acetone. A simple coordinate search for conversion of methanol to acetone would retrieve such a document from a file that provides no context.

A number of methods have been developed to introduce context to on-line databases, enabling searches to be refined to minimized false retrieval. One of the earliest techniques is proximity searching, in which two words are required to be adjacent, or within a limited distance from each other in text. The assignment of roles to chemical substances is a method of precoordinating concepts. A substance can be identified as a reactant, as a product, and in some systems in a number of additional roles. For example, by searching for documents in which formaldehyde is a product, documents in which it is a reactant, or in which it undergoes no reaction, are thus eliminated.

Another source of context comes from links between index concepts. In a database that describes chemical compounds in terms of their fragments, it is important that those fragments are tied together, and that the fragments of compound A are tied separately from the fragments of compound B. The IFI CLAIMS Patent Service's Uniterm Index has a simple method of fragmenting chemical structures, but one of its shortcomings is the fact that there is no linking of the fragments. Chemical fragments are linked in IFIs Comprehensive Data Base, and this is one of the reasons for CDBs ability to outperform Uniterm. Linking logic has enabled the articulated index terms of CA to come back into their own and restore their original context. One of the most interesting developments in on-line database indexing has been the introduction of three levels of linking in the revised polymer indexing system introduced by Derwent late in 1993. This method improves the ability of the searcher to look in an overall system for subsystems that bear a given relationship to each other.

The complex Boolean, proximity and linking technologies are available on the major search services, but such capabilities are limited on Internet using search engines available in the first years of the twenty-first century. The simplest Internet search engines can accept only a single line of search terms with Boolean connectors. More advanced search engines allow the use of proximity operators within a search or selection of search terms from different data fields, and a few allow simple Boolean combinations of answer sets. But none is capable of extended combinations of search terms, crossfile searching or linking of the type that make late twentieth century commercial search services so effective for searching patent databases.

5.2. Subject-Based Retrieval Parameters. There are numerous means by which the subject content of a patent can be expressed, and which a searcher can use in developing a search strategy. Different databases offer differing subsets of these means. Effective strategies should in general not be limited to a single type of retrieval parameter; rather, they should be built from different parameters and modified as needed to provide the strategy best fitted to the subject at hand.

Patent titles are usually short, and sometimes extremely uninformative. Patents in the latter category include titles, eg, Chemical Product and Process.

Derwent rewrites titles to make them more informative, and IFI CLAIMS augments many titles. Words included in patent titles normally are highly relevant, and title terms are thus useful in providing focused, though incomplete, retrieval.

Because a well-written abstract highlights the most important concepts in a document, words in abstracts can be highly valuable retrieval terms; however, abstracts can vary greatly in format and quality. Author abstracts on patent front pages thus run the gamut from the highly informative to the barren. Some patent abstracts can get so tied up with the legalese of patent claims or the graphical aspects of chemical structures and drawings that they become nearly useless for searching. Abstracts in CA tend to be useful for searching purposes, but CAS makes those abstracts available for search on-line only in the STN version of the CA database.

The text of patent claims is especially important for infringement searching. Many databases make available the complete claims of U.S. patents covering a period exceeding the life span of U.S. patents. Claims for European, French, United Kingdom and German patents, as well as PCT applications, are available on-line. Claims are generally searchable as individual field in full text patent databases.

The full text of patent specifications is an intriguing retrieval tool. At its best, it enables the location of the tiniest detail in patent disclosures, details that can easily escape the attention of the document analysts who abstract and index patents. At its worst, it provides discussions of prior art or alternative procedures that have no relation to the invention at hand. Full-text databases that enable a search to be limited to portions of the patent text can help improve the quality of full-text searches. Nevertheless, there are concepts embodied in drawings or structural diagrams that cannot be expressed by natural language search of full text. Further, there are aspects of rigorous and less-than-rigorous chemical nomenclature that present considerable challenges to the search of natural language text. A major pitfall in full text patent searching is that patents are published in many languages. A comprehensive full text search would require search strategies in every language represented in a database. European patents are in English, French, or German; PCT applications can be in English, French, German, Spanish, Russian, Japanese, or Chinese. Some search engines make only English text searchable, and even those that allow searching in French, German, and/or Spanish ignore the text of Russian, Japanese, and Chinese language documents. For that reason, is is not possible to perform a comprehensive search on the basis of full text.

Controlled indexing can help overcome the vagaries of free text. The structures of indexing languages can differ sharply, and thus have a substantial effect on retrieval techniques. For example, the indexing language of the Ei EnCompassPat databases is hierarchical. Terms have an interrelationship with broader, narrower, or related terms. When a document is indexed, it is indexed at the most specific level, and every indexing term generates all broader terms in the hierarchy, as well as selected related terms. Thus a search can be carried out as narrowly or as broadly as desired. The reactions of methane or of benzene can be sought specifically; consideration of the reactions of aliphatic hydrocarbons can be done in the confidence that all documents indexed for the reactions

of methane would be retrieved. All documents on the production of hydrocarbons can be looked for with the assurance that everything indexed for making either methane or benzene would be retrieved.

A different situation pertains to databases such as CA or the indexed CLAIMS files. In CA, there are generic and specific terms, but a broad generic term cannot normally be searched with confidence that all of the specifics that fall into the class will be retrieved. It is necessary to build up groups of homologues and synonyms in a search strategy, although the introduction of polymer class terms and the POLYLINK command in the CAS Registry is a great help in carrying out broad searches of polymer information. Groups of homologues and synonyms must in general be constructed in searching the Uniterm database as well; the IFI thesaurus provides helpful listings of related terms that are useful in preparing search strategy, and the company has added to the database a number of collection terms, such as zeolites and addition polymers, that are now posted to the index whenever a specific, narrower index term is used, and that have been back-posted to the file as an aid in imposing broad criteria on searches.

Patent classification codes are another subject-search parameter available in most patent databases. The IPC codes are usually present and U.S. codes exist in a number of files; in the case of Patent Abstracts of Japan, Japanese codes too are available. European Patent Classification codes are becoming more available, as database producers have begun to integrate updated classification data from the EPO into databases. It is possible to mimic a hand search by limiting operations to references falling within one class or group of classes. Although such strategies can in some instances be justified, it is usually wiser to treat class codes as just one of the various subject parameters that make up a search strategy.

5.3. Structure Searching. Fragmentation systems have been the traditional means for indexing and searching generic and Markush chemical structures in patents. Derwent's FARMDOC-AGDOC-CHEMDOC code is such a system, as are the systems used in CLAIMS-Uniterm and -CDB and in Ei EnComPassPat, but there are important differences among the systems. The Derwent system is geared only to products, not starting materials or other reactants, and thus does not include a system of roles. Fragments of a molecule are linked together to distinguish them from fragments of other substances in the same document. Uniterm has a rudimentary fragmentation system with no linking capability, so that all functional groups in all substances in the patent are thrown into the same mix. The CDB fragment system is highly detailed, and features both links to isolate the components of individual compounds and roles to denote their use. Frequently appearing chemicals that have their own descriptors are not fragmented, however, so that a search for all fluorinated alcohols must include not only the fragments for fluorine and alcohol, but also the terms for any individual fluoroalcohols contained in the indexing vocabulary. The EnCompassPat fragmentation system is less detailed than either Derwent's or CLAIMS-CDBs, but all indexed substances are fragmented, so that a purely generic search can safely be run for the fluoroalcohols without having to specify individual compounds. Another valuable EnCompassPat feature is their so-called template system of inputting the indexing for Markush formulations,

which aids in the generation of multiple linked-term sets and avoids the overcoding that produces false retrieval (31).

Fragmentation systems, useful as they are, describe molecular structures incompletely. Topological indexing systems, typified by the CAS Registry, are used to identify unambiguously each of the >26 million substances covered in CA, and can be searched for specific complete structures as well as for substructures. With the advent of the MARPAT system in 1988 the CAS began to handle generic and Markush structures as well (32,33). In the meantime, Derwent created the WPIM Markush database to deal with both exact and inexact structures in Sections B, C, and E of CPI, and INPI created PHARMSEARCH. Derwent and INPI subsequently centralized Markush DARC indexing combining the structure search databases into the Merged Markush Service. The Registry and MARPAT are searched by Messenger software on the STN system, whereas Questel-Orbit's Merged Markush Service uses Markush DARC (34–37). DIALOG also has a Registry dictionary file, CHEMNAME, which enables the compounds in the CAS Registry to be searched by a combination of parameters such as name fragments, molecular formulas, and ring system identifiers. Dictionary searching may lack some of the power of the full Registry database but it is a highly useful technique in its own right, and can be used in combination with topological searching in the STN version of the Registry.

Proteins and nucleic acids present special problems for structure searching in that they contain a small group of repeating subunits but few variable groups. Sequences of polypeptides and nucleic acids resisted efforts to search them by fragmentation or topological methods until the late 1980s, when Derwent and IntelliGenetics devised GENESEQ and CAS adapted a system for sequence searching within the Registry (38).

6. Cross-File and Multifile Techniques

Databases differ in their strengths and weaknesses, as well as in their focus. As a result, duplicate searches carried out on different databases generally produce different results. This has been demonstrated in comparative studies of retrieval results for a group of patent databases (39,40). Participants in one study (39) made an effort to develop optimal search strategies in each database tested, yet in no instance did one file produce perfect retrieval. Both investigations found that results from the various databases complemented each other. As a result, searchers are counseled to use multiple databases whenever possible. There is no pat answer to the question of how many files to use or which files to use; however, more files mean more expenditure, and searchers must develop their own cost–benefit relationship.

It is especially valuable to be able to bring elements of one database to bear on another. This technique is known as cross-file searching. Cross-file searching had its origins in the early days of CAS Registry dictionary files on the ORBIT and DIALOG systems. Searches of those files produced lists of compound registry numbers, which had to be rekeyed for searching in the versions of the CA database that these hosts offered. ORBIT developed a technique that enabled the output of a search to be obtained directly as search terms, formatted for

use in another database. All major search services now have commands that extract search terms from an answer set for searching in another database. ORBITs PRINT SELECT software was followed by DIALOGs MAP software which provided an alternative method for carrying out cross-file operations. Each of the cross-file systems has its own special characteristics. For example, STNs TRANSFER and Questel-Orbit's MEM software enable selection of only those terms meeting criteria established by the searcher, such as only the U.S. patents in a set.

Cross-file techniques permit searchers to combine the approaches and capabilities of different databases and achieve a synergistic result (41). Thus, the EI EnCompassPat database features a vocabulary that contains many specialized terms from the petroleum refining industry and has a rudimentary chemical fragmentation code. On the other hand, WPI has a fragmentation code that is much more precise. A set of candidate references can be developed on EI EnCompassPat by searching concepts, then passed against the WPI fragmentation code using cross-file techniques, to produce a search more precise than could be done on either database alone. An example of how references can be lost when one database focuses on just one of two retrieval parameters and a second database focuses on just the other parameter has been given (40). This type of situation can often be remedied by cross-file techniques.

Because API indexing was done from Derwent abstracts, and APIPAT and WPI references were cross-referenced to each other, the inherent ties between APIPAT and Derwent references along with experience in cross-file searching between APIPAT and WPI on theORBIT system, led Derwent, API, and ORBIT to undertake the physical merger of APIPAT with WPI (23). This enabled the development of search strategies by simply combining API and Derwent parameters, and confirmed the value of searches that combine the viewpoints of different databases. The desirability of a master patent database combining the features of all of the principal patent files has been expressed (39,42).

Multifile searching differs from cross-file searching in that it permits a single strategy to be brought to bear on more than one file at the same time, but the individual files are searched independently without interaction. Major on-line hosts in permit some form of multifile searching, which is exemplified by DIALOGs OneSearch. An advantage of multifile searching is that it is possible to create and use in one step a single strategy; a disadvantage is that the single strategy may not be optimum for all of the files used.

An important aspect of multifile searching is the ability of the search engine to identify duplicate records and group or remove duplicate records. This simplifies review of retrieved records and saves the cost of printing multiple references to the same patent or patent family. The absence of duplicate recognition capabilities is a major shortcoming of most Internet patent search engines.

6.1. Term Extraction and Analysis Software. Closely allied to the software used in cross-file searching is the software that extracts terms and provides statistical analysis of their occurrence within the set being analyzed. For example, a searcher may carry out a state-of-the-art search and obtain a listing of the patentees represented, ranked by the number of patents for each. The ORBITs GET software, originally used at Pergamon Infoline, a predecessor company, was the pioneer and was adapted to Questel-Orbit when the systems

merged; other systems include DIALOGs RANK, STNs SELECT, and Questel-Orbit's MEMSORT. The STN also developed a more sophisticated offshoot of the SELECT command called TRANSFER, which temporarily exits a file and revisits a file searched earlier to extract search terms from answer sets. Both RANK and SELECT enable the searcher to manipulate intermediate results to refine the product. Thus, if the output of some patentees is fragmented in the patentee ranking because of variations in spelling, such variations can be combined and the data rerun without incurring additional charges for the operation. Software of this type has become increasingly popular among searchers who need to analyze statistically various aspects of search results. Another use for term extraction software is the Family command on STN and Questel-Orbit that extracts priority application data from a patent record and searches the database for additional records sharing the same priorities, resulting in the creation of an extended patent family. The CASLINK, a software feature from STN, carries out searches of several structure files, Registry, MARPAT, and MARPAT Previews, collects the results, runs these against CA bibliographic databases, and identifies and eliminates duplicate records (43). More and more systems are being developed to simplify mechanical operations required for a searcher, and more capability is being developed to identify and, if desired, eliminate duplicate records.

7. Patent Databases

7.1. Derwent World Patents Index (WPI) and WPI Markush. Derwent's in-depth documentation services began in 1963 with the FARMDOC service for pharmaceuticals, followed by AGDOC for agricultural chemicals. PLASDOC covering polymers began in 1966. Both FARMDOC and AGDOC featured a chemical fragmentation code searched on IBM punch cards or corresponding computer tapes. A fragmentation code developed by the U.S. PTO was adopted by Derwent to handle steroid molecules. PLASDOC had its own punch card code. The code systems were severely limited because of the restrictions placed by the 960 positions on the punch card; there was considerable grouping of concepts that might better have been separated, as well as overcoding of alternatives on the same card rather than on separate cards. These limitations produced false retrieval, but such false retrieval was not a severe problem in the early days of the system, which contained modest numbers of references in the databases. Besides the punch code system, these early databases featured a manual code system for the manual searching of classified sets of documentation abstracts on search cards.

In 1970, Derwent extended its coverage to all aspects of chemistry, in the *Central Patents Index*, later renamed the *Chemical Patents Index* (CPI). The general chemical section of CPI, called CHEMDOC, was coded by a slightly modified version of the punch code system from FARMDOC and AGDOC, and manual code systems were developed for all of the nine new CPI sections. The increase in false retrieval from the growing file provided the driving force for code improvements in 1970, 1972, and 1981. Nonchemical patents were added by Derwent in 1974, expanding the total to the *World Patents Index*. Manual

codes were not at first created for the nonchemical parts of the system, but were later added for electrical patents.

With the advent of on-line searching in the 1970s, the Derwent file was one of the first to go on-line. It had subject retrieval capability by the manual and punch code systems, title terms, IPC, and broad subject groupings called Derwent classes, whose primary function had been to allocate patents to appropriate segments of the Derwent system. By 1981, abstracts were added to the database, after which abstracts for the entire back-file were added. Freed from the tyranny of the IBM punch card, Derwent gradually added retrieval parameters such as inventors and multiple patentees for jointly held patents, and strengthened the capabilities of both chemical and polymer retrieval systems. A limited number of specific chemical compounds had been directly searchable since 1981. Topological structural indexing in the WPIM file, permitting structural search of Markush structures by the Markush DARC system and direct searching for additional specific structures, was added for references from 1987 onward, and was eventually merged with the Markush DARC structure records for the INPI PharmSearch database to form the Merged Markush Service, to which earlier patent families are gradually being added. Polymer indexing was greatly enhanced with the system introduced in 1993, which featured a greatly increased indexing vocabulary and the unique capability to link data elements on three distinct levels.

An important aspect of Derwent's treatment of bibliographic data has been a standardized method for registering data elements such as patent, priority, and application numbers. Issuing countries vary in the way they assign such numbers. The number of digits can vary, and digits for the filing or publication year may precede or follow the serial number. The INPADOC standards have in many instances preserved these inconsistencies, so that searchers may be uncertain as to how to enter data in a search. Derwent standardized the presentation of these data elements, in part because of the constraints created by computer programs used before 1992, which required a fixed field length for each element. The resulting system did in some instances dictate the elision of a letter or digit, but once understood it was totally self-consistent. The Derwent format became the standard for searching patent data across the Questel-Orbit system, and is an optional standard on STN. Newer computer programs permitted Derwent to restore the elided characters in 1992, and the predictability of the system has been disrupted by changes in formats introduced by patent offices after the turn of the twenty-first century.

The WPI database offers a wide range of bibliographic and subject-based access points. Indexing is deepest in the four CPI sections having special systems, ie, FARMDOC, AGDOC, CHEMDOC, and PLASDOC. Among the first three, the origins of the system in the product-oriented pharmaceutical industry have produced a system that works less well in petrochemicals, eg, where the focus is frequently on starting materials which are often not indexed in the system. But with its combination of multiple access capabilities, detailed documentation abstracts, extensive patent family information, as well as archival and document delivery capabilities, Derwent is indisputably the overall leader in providing patent information.

7.2. Chemical Abstracts and CAS Registry. CAS is especially notable for the thoroughness and high quality of its products. The CAS Registry

system does a superb job of identifying any chemical that is either involved in new chemistry with hard data, or, since about 1980, specifically claimed in a patent. The MARPAT database has also led CAS to identify the perhaps nonexistent but prophetic substances covered by Markush claims in patents.

In its earliest years, the printed *Chemical Abstracts* provided lengthy abstracts that could often serve as surrogates for original documents. Derwent still does this in some instances, but CAS on the other hand strives to be a pointer to original documents. The volume of nonpatent and patent literature covered in CA dictates that complete abstracts are impractical, although CA indexing records can contain a wealth of detail not hinted at in abstracts.

Given the value of abstracts and chemical structures as searching tools, it is the full advantages of the CA database are available only in the STN version of the file. The mounting of the WPI database on the STN system provides powerful cross-file search capabilities between the fragmentation-coded references in the WPI system and the specifically registered references in CA. Many databases on the STN system, including the USPATFULL file, have been enhanced by the addition of CAS Registry Numbers.

7.3. INPADOC and esp@cenet. If WPI is the preeminent multipurpose file, INPADOC is the preeminent bibliographic file. Its family information is more complete than that of any other database, although it has less intellectual family data than WPI. The legal status data is the most extensive available, and has expanded from a few countries in the early 1990s to >40 countries by 2005. The subject searching capability of the database was sharply limited to only original title words (many in a variety of languages) and IPCs; some English language versions of original abstracts were added, and in appropriate circumstances INPADOC can be used effectively for subject searching. For example, if an organization located in Asia or South America has recently obtained a patent in a given area of technology, and the patent publication is so new that it has not yet reached the WPI database or is from a country not covered by Derwent, an INPADOC search can be quite fruitful.

The esp@cenet databases, available directly on the EPO and member countries' computers is unique in providing free access to bibliographic information and patent images for patents from around the world. Esp@cenet was designed for use by individual small and medium sized institutions, and lacks expensive indexing and software features common in commercial databases and search services.

7.4. CLAIMS Databases. The IFI CLAIMS offers a family of files under the CLAIMS rubric. At the heart are the three basic patent files, CLAIMS-Biblio, -Uniterm, and -CDB. The Biblio file presents front page and claims language from the patents as issued, with no added indexing. IFI does add value in a number of ways, however, which include standardized assignee names, class codes for reclassified patents, flags for reassigned and expired patents, citation counts, and searchable two-dimensional depictions of chemical structures in the claims. The Uniterm version includes indexing; it is substantial for concepts but limited for chemical structures. For subscribers, the Comprehensive Data Base has, among other things, a highly detailed structural fragmentation system and a unique system of roles applied to the indexing of polymers. The CDB system outstrips the Uniterm system by a substantial margin, and subscribers who rely on

it wish that there were more subscribers so that the cost of indexing could be more easily supported. Besides the three patent files there are a number of auxiliary files. Reexamination, reassignment, and expiration data as well as a broad range of other status change information are included in the CLAIMS Current Legal Status file, and the CLAIMS-Citation files on DIALOG are the only current source for searchable citation information on pre-1970 patents. The CLAIMS Compound Registry is an aid in searching the indexed files for specific compounds.

The time span covered by CLAIMS is unique: chemical patents from 1950, nonchemical patents from 1963. The bibliographic information for pre-1970 patents is unfortunately replete with errors, especially with respect to inventor names. On the other hand, IFI has done an admirable job of standardizing patent assignee names and correcting discrepancies and errors in the originals.

7.5. Full-Text Patent Databases. The LEXPAT database on the LEXIS-NEXIS system, the first commercially available full-text patent file, receives its greatest use from patent attorneys and has been relatively unused by other patent information specialists. This may be attributed to search software that is quite different from the type familiar to information specialists, no matter what their preferred host system. This situation has changed with the release of full text databases on Dialog, Questel-Orbit and STN, all searchable by familiar Boolean techniques and featuring greater capability for searching selected portions of patent specifications. Full patent specifications generally contain much information that is not relevant to the invention itself, and searches of full-text files, unless they are carefully framed, can produce prodigious amounts of unwanted and irrelevant answers. On the other hand, full-text searches are unique in their ability to locate the most minute passing disclosure, the type of information that can be utterly inaccessible by any other means. DIALOGs PATFULL has been enhanced by some controlled data elements from CLAIMS. The STNs USPATFULL has also been enhanced by CAS indexing of the U.S. patent or its equivalent. In addition, USPATFULL contains a thesaurus of the U.S. Manual of Classification. Special file enhancements of this sort have proliferated as a result of competition among database producers and hosts. An important feature of LEXPAT is that postissuance changes such as reassignments and corrections are incorporated directly to the file; there is no need to look for them in a second place. Full text databases are available over the Internet from some patent offices and in the subscription database services Delphion, Micropatent and Aureka. Full text patent databases for European, PCT, United Kingdom, French and German documents have become as widely available as those for U.S. patents and applications, in large part because the text from non-electronic publications has been scanned and digitized by Lexis-Nexis Univentio, Micropatent and the EPO. Micropatent and its sister service Aureka are allow searching of the full text of U.S. patents from the beginning of the modern patent system in 1836.

7.6. Other Individual Country Databases and Auxiliary Files. The U.S. patent files on Questel-Orbit, Dialog, STN, Micropatent, Aureka and Delphion, are similar in their contents to the CLAIMS-Bibliographic files, including all the front page information and the full claims language. These files do not include the two-dimensional structures provided by IFI, nor do they have IFIs

standardization of assignee names. Full text is available in many of those databases, but may be in a separate database. Searching the full text option on Delphion, for example, does not retrieve terms present only on the front page.

Several auxiliary files complement the various basic U.S. patent databases. USCLASS on Questel-Orbit coordinates updated patent classes and patent numbers going back to the beginning of the U.S. patent system. CLAIMS/Reference on DIALOG (File 124) and IFIREF on STN include the text of the U.S. classification system, along with the IFI indexing vocabulary; CLAIMS-CLASS on Questel-Orbit has just the classification text, not the indexing vocabulary. The CLAIMS-Registry shows the fragmentation of compounds specifically indexed in the IFI system, and a generic search of this file produces a search list of specific compounds to augment a generic fragment search. The CLAIMS vocabularies are available to subscribers on the IFI CLAIMS website and in a software package, IFI Ref, that generates search strategies for retrieving both specific and generic compounds. The Patent Status File from Thomson Derwent collects a wide range of post-issuance events in the life of U.S. patents, including corrections and reassignments, but is less complete than the CLAIMS Current Legal Status file. LITALERT, from Thomson Derwent, provides information on U.S. patent and trademark litigation.

Files covering the patent output of individual patent offices have proliferated. They include EPAT and for the EPO, PCTPAT and for the PCT, PATDPA and for Germany, PATDD for the former GDR, FPAT for France, as well as ITALPAT, JAPIO, GBFull for the United Kingdom, and Korean, Russian, and Chinese Patent Abstracts. Some of these have unique features and many contain images from patents. However, ITALPAT, probably the least informative of all patent databases, includes only application numbers, assignees, inventors, and titles, without even patent numbers.

JAPIO provides abstracts based in particular on patent claims, and can help to clarify uncertainties with Japanese abstracts from Derwent and/or CA. PATOLIS, in Japanese, is a unique source of Japanese legal status information (44). EPIDOS staff carry out PATOLIS searches on request; for those with sufficient need to search the PATOLIS database, access is available for a monthly fee to the PATOLIS-e version of the files with an English interface to the database.

7.7. Other Databases with Patent Information. The EI EnCompassPat database has been discussed, as have the unique capabilities of the merged WPI-EI EnCompassPat file on Questel-Orbit. Many other databases contain substantial amounts of patent information, notable among them are Derwent's *Biotechnology Abstracts*, the TULSA database (petroleum exploration and production), several specialized pharmaceutical files, PAPERCHEM, and METADEX. A very complete listing of databases containing patent information circa 1990 (45) summarizes the patent content of the databases available at that time.

8. Archiving and Document Delivery

Many organizations have traditionally maintained in-house collections of patent specifications in areas of interest. Microforms and, more recently, CD-ROMs took the place of paper copy as the volume of the patent literature grew and in-house

collections of patent specifications have now largely been discontinued in favor of on-demand delivery of patent specification over the Internet. National and regional patent offices make copies of patents available at no charge over the Internet, although many of the offices have deferred to commercial suppliers by making patent specifications available as single page images. Some commercial suppliers obtain patent copies by downloading patents from esp@cenet, the U.S. PTO or other websites and combining the patents into a single PDF file. Others, notably Thomson Scientific's MicroPatent, allow direct downloading of PDF files of patent documents from a great many countries for a moderate price or as part of a subscription that includes access to a searchable database.

Users of patent information occasionally need patents that are not available from the extensive patent copy collections available electronically and must order copies from other sources. National patent offices such as the U.S. PTO have been important suppliers of copies, typically for their own country and other countries as well. In the United States, there are more than 70 patent depository libraries that serve as regional patent information resources. In addition, a number of other organizations supply patent copies, including Thomson Scientific, CAS, DIALOG, EPIDOS, and others. Electronic means of document delivery iscommon; telefacsimile (FAX) and email are used more often that mail.

9. Other Technological Initiatives

The ready availability of searchable text from patents produced electronically by the issuing patent offices has created a demand for complete access to the full text of all patents. To obtain the text of patent documents published only on paper, it is necessary to scan the documents and use optical character recognition software to convert the printed text into computer readable text. Optical character recognition does not reproduce the printed text perfectly; errors can be introduced by poorly converted fonts, unavailable field identifiers, old or annotated original documents, non-alphanumerical characters such as Greek letters and superscripts, and innapropriate spell checking. The first attempts at large scale digitization targeted United States patents produced before electronic document publication was introduced in the early 1970s and Patent Cooperation Treaty applications, which continue to be published only in print. Lexis-Nexis Univentio has created files of digitized patent documents from many other countries and provides full text data under license to commercial search services. The Lexis-Nexis Total Patent service provides subscribers with full text from 22 patent issuing authorities in addition to bibliographic information from up to 100 offices. Univentio has promised to re-scan the documents in the databases on a regular basis with continually improving technology. In addition, it provides fully searchable English machine-assisted translations for many of the non-English documents. Machine-assisted translation is increasingly available for Japanese patent documents, but had not become widely available for other languages as of 2005. For Japanese Kokai published electronically since 1993 nearly instantaneous electronic translations are available from the Japan Intellectual Property Digital Library at no cost and from PATOLIS and the US-based Paterra service for a moderate fee. The quality of the translations varies widely among

the services and among technologies, depending on the quality of the underlying lexicons. Paterra's translations for chemical technologies are exceptionally accurate, reflecting the chemical training and experience of the proprietor.

Strategies developed using Boolean logic have been central to the searching of on-line databases, but during the 1990s it was suggested that effective searches could be carried out with natural language input by using systems based on artificial intelligence. Some of the simpler tools are relevance ranking algorithms embedded in commercial online search services and Internet search engines; more complex tools can be very expensive. Simple relevance ranking software simply assigns a numerical value to each record and lists them with the highest ranked records at the front of the list, a system familiar to users of the search engines on the World Wide Web. Relevance ranking can simply count the number of times search terms occur in an the records in an answer set, or it can factor in their proximity to each other, their position in the record and the total number of terms in the record. DIALOGs TARGET software and STNs FOCUS are aimed not particularly at patents, but at any database containing substantial amounts of text, especially full-text files. Full text is desirable because the system is based in part on term-frequency counts, but the number of records analyzed and reported is somewhat limited. More complex linguistic tools factor in the frequency a search term occurs in the language as a whole. One example, the Themescape mapping tool provided with the Aurigen search service purports to identify concepts in a large set of patent records rather than simple words. The software generates a topological map of the concepts it identifies with related or cooccuring concepts in proximity to each other and highly posted terms shown as peaks separated by valleys. The linguistic concept generator makes the assumption that very rare or very frequently occuring terms are irrelevant to the analysis, an assumption that is less likely to be correct for the vocabulary in technical documents like patents than in conversational language. Considering the amount of information conveyed by chemical structures and drawings and the essential function of claims in patents, the relevance of relevance ranking in patent databases is highly suspect (46).

Data mining software takes a different approach toward patent analysis. Using large sets of records from fielded databases, the software generates cooccurrence matrices, trends over time, and distributions of patentees, inventors or technology-related codes. These tools must be customized to recognized the fields and formats in a particular database and search service. Simple data analysis tools are built into the major commercial search services, and more powerful tool sets are available. Tools can be dedicated to a single search service, as is STNs Anavist, or can be programmed to recognize fields from any database of interest to the analyst, as can Vantage Point, a software tool produced by Search Technology, Inc. Data mining and text mining have been termed "patinformatics" (47), and can be useful in providing a high level overview of trends in technology quite unlike the specific views provided by traditional patent searches.

BIBLIOGRAPHY

"Patent (Literature)" in *ECT* 1st ed., Vol. 9, pp. 890–897; in *ECT* 2nd ed., Vol. 14, pp. 583–635, by E. S. Turner, Bell Telephone Laboratories, Inc.; in *ECT* 3rd ed., Vol. 16, pp. 889–945, by J. W. Lotz, IFI/Plenum Data Co.; "Patents, Literature" in *ECT* 4th ed., Vol. 18, pp.102–156, by E. S. Simmons, Hoechst Marion Roussel, Inc., and S. M. Kaback, Exxon Research and Engineering Co.; "Patents, Literature" in *ECT* (online), posting date: December 4, 2000, by E. S. Simmons, Hoechst Marion Roussel, Inc., and S. M. Kaback, Exxon Research and Engineering Co.

CITED REFERENCES

1. *Derwent Guide to Patent Expiries*, Edition 9, 2004. Thomson Derwent, London.
2. *Manual of Patent Examining Procedure*, Section 901.04 [8th ed., Revision 14, May 2004], U.S. Department of Commerce, Patent and Trademark Office, Washington, D.C. http://www.uspto.gov/web/offices/pac/mpep/documents/0900_901_04.htm.
3. World International Property Office Handbook of Industrial Property Information and Documentation, Standard St. 16, pp. 3.16.1-15, June 2001 http://www.wipo.int/scit/en/standards/pdf/03-16-01.pdf.
4. WO 95/00525.
5. A. H. Berks, *World Pat. Inf.* **23**(19): 5 (2001).
6. A. H. Berks, Markush Structures in Patents, in P. V. R. Schleyer and co-workers, eds., *The Encyclopedia of Computational Chemistry*, John Wiley & Sons, Inc., Chichester, 1998, Vol. 3, pp. 1552–1559.
7. WIPO, *International Patent Classification*. 7th ed., WIPO Publication 560/7th, World Intellectual property Organization, 1999.
8. I. Schellner, *World Pat. Inf.* **24**, 197 (2002).
9. D. T. Dickens, *World Pat. Inf.* **16**, 28 (1994).
10. *Manual of Classification*, U.S. Department of Commerce, Patent and Trademark Office, Washington, D.C.; revised regularly.
11. *Patent Classification Definitions*, U.S. Department of Commerce, Patent and Trademark Office, Washington, D.C.; revised regularly.
12. S. R. Adams, *World Pat. Inf.* **23**, 15 (2001).
13. S. M. Kaback, in H. F. Mark and co-workers, eds., *Encyclopedia of Polymer Science and Engineering*, 2nd ed., Vol. 10, John Wiley & Sons, Inc., New York, 1987, pp. 787–801.
14. S. M. Kaback, *J. Chem. Inf. Comput. Sci.* **31**, 439 (1991).
15. J. A. Briggs, E. A. Ferns, and K. E. Shenton, *J. Chem. Inf. Comput. Sci.* **31**, 454 (1991).
16. S. M. Kaback, *World Pat. Inf.* **16**, 60 (1994).
17. E. S. Simmons, *J. Chem. Inf. Comput. Sci.* **24**, 10 (1984).
18. M. Z. Balent and J. M. Emberger, *J. Chem. Inf. Comput. Sci.* **15**, 100 (1975).
19. N. Lambert, *Database*, **10**(6) 46 (1987).
20. M. P. O'Hara and C. Pagis, *J. Chem. Inf. Comput. Sci.* **31**, 59 (1991).
21. E. H. Brenner, *World Pat. Inf.* **9**, 27 (1987).
22. A. Girard, K. Landsberg, and S. M. Kaback, *Database*, **1**(2) 46 (1978).
23. S. M. Kaback, *Database*, **13**(2) 22 (1990).
24. E. S. Simmons, *J. Chem. Inf. Comput. Sci.* **25**, 379 (1985).
25. E. S. Simmons, *Database*, **18**(3) 28 (1995).
26. M. P. Carpenter, F. Narin, and P. Woolf, *World Pat. Inf.* **3**, 160 (1981).

27. M. E. Mogee and R. G. Kolar, *World Pat. Inf.* **14**, 237 (1992).
28. S. M. Kaback, *World Pat. Inf.* **6**, 80 (1984).
29. E. S. Simmons and S. M. Kaback, *World Pat. Inf.* **10**, 204 (1988).
30. E. S. Simmons and N. Lambert, in H. Collier, ed., *Recent Advances in Chemical Information*, CRC Press, Inc., Boca Raton, Fla., 1992, pp. 33–78.
31. S. M. Kaback, *World Pat. Inf.* **16**(2) 1994.
32. W. Fisanick, *J. Chem. Inf. Comput. Sci.* **30**, 145 (1990).
33. T. Ebe, K. H. Sanderson, and P. S. Wilson, *J. Chem. Inf. Comput. Sci.* **31**, 31 (1991).
34. K. A. Cloutier, *J. Chem. Inf. Comput. Sci.* **31**, 40 (1991).
35. E. S. Simmons, *J. Chem. Inf. Comput. Sci.* **31**, 45 (1991).
36. N. R. Schmuff, *J. Chem. Inf. Comput. Sci.* **31**, 53 (1991).
37. J. M. Barnard, *J. Chem. Inf. Comput. Sci.* **31**, 64 (1991).
38. A. H. Berks, *Trends Biotech.* **12**, 352 (1994).
39. S. M. Kaback, *Online*, **7**(4) 22 (1983).
40. B.-R. Weckend, *World Pat. Inf.* **9**, 140 (1987).
41. S. M. Kaback, *Database*, **10**(5) 17 (1987).
42. M. A. Lobeck, *World Pat. Inf.* **16**, 14 (1994).
43. R. E. Buntrock, *Database* **17**(3) 116 (1994).
44. G. Vacek, *World Pat. Inf.* **16**, 41 (1994).
45. J. F. Sibley, *Online Patents, Trade Marks, and Service Marks Databases*, ASLIB, London, 1991.
46. E. S. Simmons, *J. Chem. Inf. Comp. Sci.* **39**, 453 (1999).
47. A. J. Trippe, *World Pat. Inf.* **25**, 211 (2003).
48. A. J. Trippe, *Searcher.* **10**(9) 28 (2002).

GENERAL REFERENCES

Manual of Patent Examining Procedure. U.S. Department of Commerce, Patent and Trademark Office, Washington, D.C.; revised periodically. Published in print and on the uspto.gov website.

M. A. Epstein, *Modern Intellectual Property*, 3rd ed., Aspen Law and Business, Englewood Cliffs, N.J., 1995.

B. A. Amernick, *Patent Law for the Nonlawyer: A Guide for the Engineer, Technologist, and Manager*, 2nd ed., Van Nostrand Reinhold, Co., Inc., New York, 1991.

A. Reverdin and F. Schlaepfer, eds., *Katzarov's Manual on Industrial Property All Over the World*, Katzarov SA, Geneva, 9th ed., 1981; 11th rev., 1991.

J. Maynard and H. Peters, *Understanding Chemical Patents*, 2nd ed., American Chemical Society, Washington, D.C., 1991.

B. M. Rimmer, *International Guide to Official Industrial Property Publications*, 3rd ed., revised and updated by S. Van Dulken, British Library, London, 1992.

P. D. Rosenberg, *Patent Law Basics*, Clark Boardman Callaghan, Deerfield, Ill., 1992; revised annually.

E. S. Simmons, in C. J. Armstrong and J. A. Large, eds., *Manual of Online Search Strategies*, 3rd ed., (Gower, Aldershot, U.K.), 2001, Vol. 2, pp. 23–140.

S. Ashpitel, D. Newton, and S. Van Dulken, ed., *Introduction to Patents Information*, 4th ed., British Library Publishing, London, 2002.

W. Warr and C. Suhr, *Chemical Information Management*, VCH Publishers, New York, 1992.

WIPO, "Glossary of Terms Concerning Industrial Property Information and Documentation," *World Pat.Inf.* **15**, 21 (1993).

H. J. Knight, Patent Strategy for Researchers and Research Managers. John Wiley & Sons Ltd., Chichester, 1996.

L. McLeland, *What Every Chemist Should Know About Patents*, 3rd ed. American Chemical Society, Washington, D.C., 2003. http://www.chemistry.org/government/patentprimer.pdf.

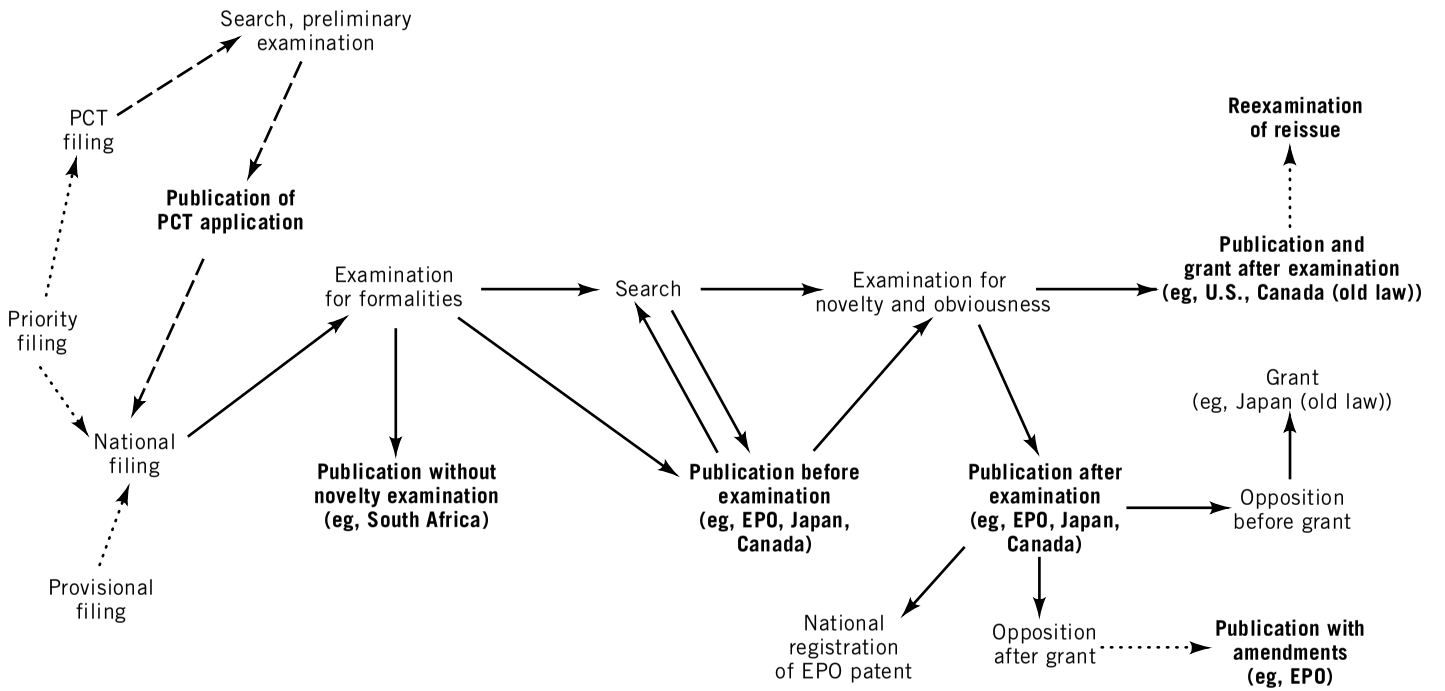


Fig. 1. Procedures for publication of patent documents. Publications are shown in boldface. Dotted lines indicate events that may take place before a national application has been filed or after a patent is granted. Dashed lines indicate events that take place only when PCT filing is chosen.

(12) **United States Patent**
Ahlers

(10) **Patent No.: US 6,835,855 B2**
(45) **Date of Patent: Dec. 28, 2004**

(54) **METAL COMPLEX CARRYING A 2-PHOSPATRICYCLO[3.3.1.1(3.7)]DECYL RADICAL AS A LIGAND IN HYDROFORMYLATION**

(75) Inventor: **Wolfgang Ahlers**, Worms (DE)

(73) Assignee: **BASF Aktiengesellschaft**, Ludwigshafen (DE)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 92 days.

(21) Appl. No.: **10/275,562**

(22) PCT Filed: **May 11, 2001**

(86) PCT No.: **PCT/EP01/05405**

§ 371 (c)(1),
(2), (4) Date: **Nov. 7, 2002**

(87) PCT Pub. No.: **WO01/85661**

PCT Pub. Date: **Nov. 15, 2001**

(65) **Prior Publication Data**

US 2003/0092935 A1 May 15, 2003

(30) **Foreign Application Priority Data**

May 12, 2000 (DE) 100 23 468

(51) **Int. Cl.**[7] **C07C 45/50**; C07F 9/00

(52) **U.S. Cl.** **568/454**; 568/452; 568/457; 556/20

(58) **Field of Search** 568/454, 457, 568/452; 556/20

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,050,531 A * 8/1962 Epstein et al.
6,037,500 A * 3/2000 Zhang

FOREIGN PATENT DOCUMENTS

GB 1 573 422 8/1980
WO 98/42717 10/1998

OTHER PUBLICATIONS

XP-001010653, Gee et al. CHem.Comm.1999,901–902.
Organometallics 2000, 19, 2791–3796 XP-002173920, Carraz et al.

* cited by examiner

Primary Examiner—Johann Richter
Assistant Examiner—Sikarl A. Witherspoon
(74) *Attorney, Agent, or Firm*—Keil & Weinkauf

(57) **ABSTRACT**

In a process for the hydroformylation of ethylenically unsaturated compounds, at least one ethylenically unsaturated compound is reacted with carbon monoxide and hydrogen in the presence of a ligand-metal complex of ruthenium, rhodium, palladium, iridium and/or platinum, where the ligand-metal complex comprises a monophosphine, monophosphinite or monophosphinamidite ligand of the formula I

(I)

where

A together with the phosphorus atom to which it is bound forms a 2-phosphatricyclo[3.3.1.1{3,7}]decyl radical in which one or more nonadjacent carbon atoms may be replaced by heteroatoms and which may be substituted, and

R' is hydrogen or an organic radical having a molecular weight of up to 20 000 bound via a carbon atom or oxygen atom or nitrogen atom.

The process is particularly useful for the hydroformylation of internal branched olefins and lower pressures and/or temperatures are required in the hydroformylation than when using other phosphorus ligands.

6 Claims, No Drawings

Fig. 2. Representative front page of a U.S. patent. The bracketed numbers are INID Codes. For example, code [54] designates the title of an invention.

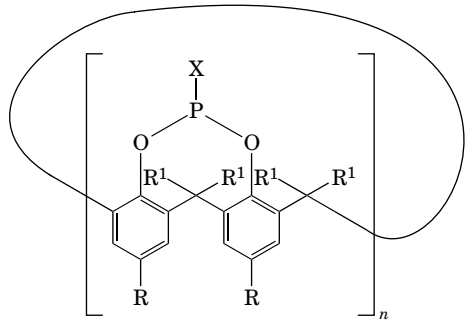


Fig. 3. Typical Markush structure where R = H, C_{1-20} alkyl, C_{1-20} alkoxy, sulfonate, or carboxylate; R^1 = H or C_{1-20} alkyl; and X is H, C_{1-20} alkyl, phenyl (Ph), or OPh, the last two of which may be substituted by up to three groups selected from C_{1-20} alkyl, C_{1-20} alkoxy, sulfonate, carboxylate, C_{1-20} alkylthio-, and/or C_{2-20} dialkylamino-(4).

C	**07**	**c**	45/00 Main group
Section			or
(chemistry)	**Class**		**45/50 Subgroup**
	(organic chemistry)	**Subclass**	
		(of acyclic compounds)	

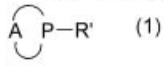
45/00 Preparation of compounds having >C=O groups bound only to carbon or hydrogen atoms; Preparation of chelates of such compounds [2]
45/49 .by reaction with carbon monoxide [3]
45/50 ..by oxo reactions [3]

Fig. 4. Example of international patent classification (structured, hierarchical), where numbers in square brackets identify edition of IPC in which class was first used. In C07c 45/50, the first four characters indicate section C (chemistry), Class 07 (organic chemistry), and subclass c (acyclic compounds); the number 45/00 indicates the preparation of compounds having carbonyl groups bound only to carbon or hydrogen atoms by any method; and 45/50 indicates such preparation by oxo reactions. Beginning with the 8th ed. in 2006, advanced level update codes will be shown in parentheses.

CLASS 568	ORGANIC COMPOUNDS
568-300	.OXYGEN CONTAINING
420	..Aldehydes
448	...Acyclic
449	....Processes
450	.....Isomerization
451	.....Hydroformylation by reacting ethylenically unsaturated compound, carbon monoxide, and gaseous hydrogen
452	.....Dimer produced
453	.....Plural stages each having hydroformylation
454	.....Group 15 (VA) element (N, P, As, Sb, or Bi) containing material utilized (eg, arsenic containing ligand utilized, etc)

Fig. 5. Example of U.S. patent classification (unstructured, hierarchical). This is one of a series of classes considered as integral parts of Class 260, following the schedule hierarchy retaining all pertinent definitions and class lines of Class 260.

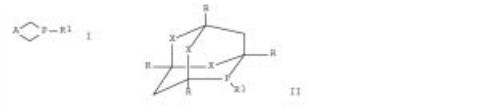
L1 ANSWER 1 OF 1 WPIX COPYRIGHT 2005 THE THOMSON CORP on STN
AN 2002-076150 [11] WPIX Full-text
DNC C2002-022966
TI Novel ligand-metal complex, useful for hydroformylation of ethylenically unsaturated compounds, comprises ruthenium, rhodium, palladium and platinum and a monophosphine, monophosphinite or monophosphinamidite ligand.
DC E11 E19 J04
IN AHLERS, W
PA (BADI) BASF AG; (AHLE-I) AHLERS W
CYC 97
PI DE----10023468 A1 20011115 (200211)* 9 C07C-045-50
WO---200185661 A1 20011115 (200211) GE C07C-045-50
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW
W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW
AU---200174027 A 20011120 (200219) C07C-045-50
EP-----1280754 A1 20030205 (200310) GE C07C-045-50
R: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI TR
US--2003092935 A1 20030515 (200335) C07C-045-49
EP-----1280754 B1 20040804 (200451) GE C07C-045-50
R: BE DE GB
DE----50103117 G 20040909 (200459) C07C-045-50
US-----6835855 B2 20041228 (200502) C07C-045-50 <--
ADT DE----10023468 A1 2000DE-1023468 20000512; WO---200185661 A1 2001WO-EP005405 20010511; AU---200174027 A 2001AU-0074027 20010511; EP-----1280754 A1 2001EP-0940459 20010511, 2001WO-EP05405 20010511; US--2003092935 A1 2001WO-EP05405 20010511, 2002US-0275562 20021107; EP-----1280754 B1 2001EP-0940459 20010511, 2001WO-EP05405 20010511; DE----50103117 G 2001DE-0503117 20010511, 2001EP-0940459 20010511, 2001WO-EP05405 20010511; US-----6835855 B2 2001WO-EP05405 20010511, 2002US-0275562 20021107
FDT AU---200174027 A Based on WO---200185661; EP-----1280754 A1 Based on WO---200185661; EP-----1280754 B1 Based on WO---200185661; DE----50103117 G Based on EP-----1280754, Based on WO---200185661; US-----6835855 B2 Based on WO---200185661
PRAI 2000DE-1023468 20000512
IC ICM C07C-045-49; C07C-045-50
ICS B01J-031-12; B01J-031-24; C07C-051-14; C07F-009-00; C07F-015-00



(1)

AB DE 10023468 A UPAB: 20020215
NOVELTY - A novel ligand-metal complex comprises ruthenium, rhodium, palladi and platinum and a monophosphine, monophosphinite or monophosphinamidite lig
DETAILED DESCRIPTION - A ligand-metal complex (I) comprises ruthenium, rhodi palladium and platinum and a monophosphine, monophosphinite or monophosphinamidite ligand of formula (1). An INDEPENDENT CLAIM is included the hydroformylation of ethylenically unsaturated compounds by reaction of a least one ethylenically unsaturated compound with carbon monoxide and hydrog in the presence of a ligand-metal complex (I). A = together with the phospho atom it is bonded to comprises a 2-phospha-tricyclo(3.3.1.1(3,7)) decyl grou optionally having non-neighboring carbon atoms substituted by oxygen atoms a optionally substituted;

Fig. 6. Representative Derwent documentation abstract. (Courtesy of Derwent Information Ltd.)



AB A process is provided for the hydroformylation of ethylenically unsatd. comp in which at least one ethylenically unsatd. compound is reacted with CO and in the presence of a metal-ligand complex of Ru, Rh, Pd, Ir and/or Pt, and i which the metal-ligand complex contains a monophosphine, monophosphinite or monophosphinamidite ligand I [A, together with the P atom to which it is bon forms an (un)substituted 2- phosphatricyclo[3.3.1.1{3,7}]decyl radical in wh ons or more nonadjacent C atoms are optionally replaced by heteroatoms; R' = or an organic radical with a mol. weight of up to 20,000 which is bonded wit C, O or N atom]. Preferably, the ligand has the general structure II (X = O NR; R = H, alkyl, cycloalkyl, halcalkyl, aryl, aralkyl; R1 = H, alkyl, cycloalkyl, aryl, aralkyl, alkoxy, cycloalkoxy, aryloxy, aralkyloxy, alkylam dialkylamino, cycloalkoxyamino, N-cycloalkyl-N-alkylamino, dicycloalkylamino arylamino, N-aryl-N-alkylamino, diarylamino, aralkylamino, N-aralkyl-N- alkylamino, bis(aralkyl)amino, acyl, carbamoyl). In an example, treating 3 g octene with 10 bar CO/H2 (1:1) at 100° in 3 g PEMe containing 0.9 mg Rh(CO)2(acac) and 52 mg 2-cyclohexyl-2-phospha-1,3,5,7-tetramethyl-6,9,10- tricyatricyclo[3.3.1.1(3,7)]decane (ligand to metal ratio = 50:1) gave 98% conversion, with 99% aldehyde selectivity. The method is particularly suita for the hydroformylation of internally branched olefins and is characterized that lower pressures and/or temps. are required for the hydroformylation tha with other P ligands.

ST hydroformylation alkene metal phosphatricyclodecene complex catalyst; synthesis gas hydroformylation alkene rhodium phosphatricyclodecene catalyst

IT Hydroformylation
(hydroformylation of alkenes in presence of phosphatricyclodecene complexes of Group VIII metals)

IT Alkenes, reactions
RL: RCT (Reactant); RACT (Reactant or reagent)
(hydroformylation of alkenes in presence of phosphatricyclodecene complexes of Group VIII metals)

IT Synthesis gas
(hydroformylation of alkenes with synthesis gas in presence of phosphatricyclodecene complexes of Group VIII metals)

IT Group VIII element complexes
RL: CAT (Catalyst use); USES (Uses)
(hydroformylation of alkenes with synthesis gas in presence of phosphatricyclodecene complexes of Group VIII metals)

IT Unsaturated compounds
RL: RCT (Reactant); RACT (Reactant or reagent)
(hydroformylation of unsatd. compds. in presence of phosphatricyclodecene complexes of Group VIII metals)

IT Hydroformylation catalysts
(phosphatricyclodecene complexes of Group VIII metals)

IT 7439-88-5D, Iridium, derivs., uses 7440-05-3D, Palladium, derivs., uses 7440-06-4D, Platinum, derivs., uses 7440-16-6D, Rhodium, derivs., uses 7440-18-8D, Ruthenium, derivs., uses
RL: CAT (Catalyst use); USES (Uses)
(hydroformylation catalyst with phosphatricyclodecene derivative ligands)

L2 ANSWER 1 OF 1 CAPLUS COPYRIGHT 2005 ACS on STN
AN 2001:833251 CAPLUS Full-text
DN 135:371863
ED Entered STN: 16 Nov 2001
TI Process for hydroformylation of alkenes with synthesis gas in presence of metal complex carrying a 2-phosphatricyclo[[3.3.1.1(3,7)]]decyl ligand
IN Ahlers, Wolfgang
PA BASF Aktiengesellschaft, Germany
SO PCT Int. Appl., 24 pp.
CODEN: PIXD2
DT Patent
LA German
IC ICM C07C-045/50
ICS B01J-031/24; C07F-015/00
CC 29-7 (Organometallic and Organometalloidal Compounds)
Section cross-reference(s): 21
FAN.CNT 1

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
PI WO2001085661	A1	20011115	2001WO-EP05405	20010511
W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
			2000DE-1023468	A 20000512
DE--10023468	A1	20011115	2000DE-1023468	20000512
EP---1280754	A1	20030205	2001EP-0940459	20010511
EP---1280754	B1	20040804		
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO, MK, CY, AL, TR				
			2000DE-1023468	A 20000512
			2001WO-EP05405	W 20010511
US2003092935	A1	20030515	2002US-0275562	20021107 <--
US---6835855	B2	20041228		
			2000DE-1023468	A 20000512
			2001WO-EP05405	W 20010511

CLASS

PATENT NO.	CLASS	PATENT FAMILY CLASSIFICATION CODES
WO 2001085661	ICM	C07C-045/50
	ICS	B01J-031/24; C07F-015/00
WO2001085661	ECLA	B01J031/18C; B01J031/24; C07C045/50; C07F015/00N6B
DE--10023468	ECLA	B01J031/18C; B01J031/24; C07C045/50; C07F015/00N6B
US2003092935	NCL	568/454.000; 556/020.000; 568/452.000; 568/457.000
	ECLA	B01J031/18C; B01J031/24; C07C045/50; C07F015/00N6B <--

OS CASREACT 135:371863; MARPAT 135:371863
GI

IT 28473-21-4P, Nonanol
RL: BYP (Byproduct); PREP (Preparation)
(hydroformylation of alkenes in presence of rhodium catalyst having phosphatricyclodecene derivative ligand)

IT 14874-82-9
RL: CAT (Catalyst use); USES (Uses)
(hydroformylation of alkenes in presence of rhodium catalyst having phosphatricyclodecene derivative ligand)

IT 111-66-0, 1-Octene 9021-92-5
RL: RCT (Reactant); RACT (Reactant or reagent)
(hydroformylation of alkenes in presence of rhodium catalyst having phosphatricyclodecene derivative ligand)

IT 124-19-6P, Nonanal
RL: SPN (Synthetic preparation); PREP (Preparation)
(hydroformylation of alkenes in presence of rhodium catalyst having phosphatricyclodecene derivative ligand)

IT 26098-25-5 97719-46-3 98404-12-7 301525-54-2 373387-34-9
RL: CAT (Catalyst use); USES (Uses)
(ligand for Group VIII metal hydroformylation catalyst)

RE.CNT 4 THERE ARE 4 CITED REFERENCES AVAILABLE FOR THIS RECORD

RE

(1) Charles-Antoine, C; ORGANOMETALLICS 2000, V19(19), P3791
(2) Gee, V; CHEM COMMUN (CAMBRIDGE) 1999, 10, P901
(3) Shell Int Research; GB---1573422 A 1980 HCAPLUS
(4) Shell Int Research; WO---9842717 A 1998 HCAPLUS

Fig. 7. Representative abstract from *Chemical Abstracts*. (Courtesy of the American Chemical Society.)

Table 1. **Recent Milestones in the Development of Primary Patent Literature**

Year	Country or authority	Significant development
1964	The Netherlands	First principal examining office to switch to universal publication and deferred examination of patent applications
1968	Germany, Federal Republic	Switch to universal publication; a huge backlog of pending cases published, often at a rate of over 1000/week, straining patent documentation services
1971	Japan	Switch to universal publication: output rose quickly to over 100,000/year; language and numbers make quality documentation a substantial problem for abstracting and indexing services
1978	European Patent Office (EPO)	Begins operation, offering a single patent covering multiple countries: increasingly supplanted patent issuing by national patent offices, some of which have closed the national patenting route; patent publications predominently in English
1978	World Intellectual Property Organization (WIPO)	Single patent application submitted to multiple countries and regional offices through the Patent Cooperation Treaty; further increased share of English-language documents as usage increased during the 1990s
1995	United States	Switch from patent term of 17-years from grant to 20-years from filing for patents filed later than 7 June 1995; introduction of provisional applications
1995	Worldwide	TRIPS provisions of General Agreement on Tariffs and Trade established the World Trade Organization and mandated liberal patent laws for member countries
2001	United States	Pregrant publication of most patent applications filed on or after 29 November 2000; patent term adjustment possible for patents filed on or after 29 May 2000
2002	United States	Paper publication of USPTO Official Gazette discontinued; official notification of patent issuance becomes electronic
2005	Worldwide	Deadline for WTO member countries to amend laws to provide patent terms of at least 20 years from filing and claims to products as well as processes
2005	European Patent Office	Paper publication of patent specifications discontinued

Table 2. **Multicountry Patent Databases**

Database	Systems	Producer	Coverage
World Patents Index	DIALOG, QUESTEL-ORBIT, STN, Delphion, Westlaw	Thomson Derwent	International; limited bibliographic data; patent families; comprehensive English language abstracts of basic and some equivalent granted patents; topological structure indexing; polymer and chemical structure indexing for subscribers; indexing for drawings
INPADOC	PATOLIS, DIALOG, QUESTEL-ORBIT (in FamPat), STN, Delphion, Minesoft	European Patent Office	International; bibliographic data; some abstracts; patent families; limited legal status data for 46 countries
FamPat, PlusPat	Questel-Orbit	Questel-Orbit	International; combines all data from Inpadoc with national patent databases on the system; patent families; European Patent Classification codes
CLAIMS	DIALOG, QUESTEL-ORBIT, STN	IFI CLAIMS Patent Service	United States; full bibliographic data; abstract and full claim text; chemical structure, general concept and patentee name coding
PATOLIS, PATOLIS-e	PATOLIS	Japan Patent Information Organization	Japan; full bibliographic data; abstract and first claim text in Japanese; status data; drawings
Patent Abstracts of Japan (PAJ)	DIALOG, QUESTEL-ORBIT, STN. Delphion, etc.	National Center for Intellectual Property Information & Training	Japan; bibliographic data and English language abstract; drawings
Chinese Patent Abstracts	DIALOG, QUESTEL-ORBIT	European Patent Office	China; bibliographic data and English language abstract
Korean Patent Abstracts	STN	Korean Institute of Patent Information	Korea; bibliographic data and English language abstract; drawings
Russian Patent Abstracts	STN	Russian Agency for Patents and Trademarks	Russia; bibliographic data and English language abstract; drawings
Derwent Patents Citation Index	DIALOG, STN	Thomson Scientific	International; examiners' and some inventors' citations of earlier references and later citing patents; Derwent title, bibliographic, and family information
CLAIMS-Citation	DIALOG	Search Check, Inc.; IFI CLAIMS Patent Service	United States; patent numbers; examiners' citations of earlier patents; later citing U.S. patents

EnCompassPat	DIALOG, STN	Elsevier Engineering Information, Inc.	International, petroleum, petrochemical; limited bibliographic data; comprehensive English language abstracts of basic patents; concepts and chemical structure coding for subscribers
World Patents Index/ EnCompassPat	QUESTEL-ORBIT	Thomson Scientific; Elsevier Engineering Information, Inc.	International; limited bibliographic data; patent families; comprehensive English language abstracts of basic patents; all World Patents Index and APIPAT indexing for subscribers
CA File, CAPlus, CA Previews	STN	Chemical Abstracts Service	International, chemistry; limited bibliographic data; comprehensive English language abstracts; deep indexing of chemical concepts; structure-searchable compound registry; chemical structure drawings
MARPAT, MARPAT Previews	STN	Chemical Abstracts Service	International, chemistry; structure-searchable Markush formulas in addition to other data searchable in the CA File, CAPlus, or CA Previews
CA Search	Data-Star, DIALOG, QUESTEL-ORBIT, etc	Chemical Abstracts Service	International, chemistry; limited bibliographic data; deep indexing of chemical concepts; compound registry searchable by name and molecular features on some systems
MDL Patent Chemistry Database	MDL Discovery-Gate, CrossFire	MDL	United States, EP, WO; organic chemical reactions, substances and substance related information; structure searchable
Journal of Synthetic Methods	STN	Thomson Scientific	International; organic chemical reactions from patents and journal literature; structure searchable
Biotechnology Abstracts	Data-Star, DIALOG, QUESTEL-ORBIT, STN	Thomson Scientific	International, biotechnology; limited bibliographic data; comprehensive English language abstracts of basic patents
GENESEQ	STN	Thomson Scientific	International, biotechnology; limited bibliographic data; polypeptide and nucleic acid sequences and related indexing

PHARM-SEARCH	Questel-Orbit	Institut National de la Propriété Industrielle	International, pharmaceutical; bibliographic data; English language abstract and pharmaceutical indexing; chemical structure drawings
Drug Patents Interna-tional	Data-Star, DIALOG, QUESTEL-ORBIT, STN	IMSWorld Publications Ltd.	International, marketed pharmaceuticals; bibliographic data; information relating to drugs covered by patents
Delphion	Internet http://www.delphion.-com/	Thomson Scientific	United States, EP, WO, German, PAJ, INPADOC, IP.com; searchable full text and/or front page information; links to family, status, and cited/citing patents for US patents; full patent images; PDF document delivery.
MicroPatent PatentWeb	Internet http://www.micropat.-com/	Thomson Scientific	United States, EP, WO, Germany, France, United Kingdom, PAJ, INPADOC; searchable full text and/or front page information; patent families; drawings; PDF document delivery.
Aureka	Internet http://www.aureka.com/	Thomson Scientific	United States, EP, WO, Germany, France, United Kingdom, PAJ; searchable full text and/or front page information; drawings; PDF document display; file sharing; complex data mining capabilities.
PatBase	Internet http://www.minesoft.-com	Minesoft	United States, EP, WO, Germany, France, United Kingdom, PAJ, INPADOC; searchable full text and/or front page information; patent families; drawings; PDF document delivery.
LexisNexis Patent Warehouse	Internet http://www.univentio.-com	LexisNexis Univentio	International.; combines all data from Inpadoc with national patent databases on the system, including O.C.R. text from less widely available countries

Table 3. **Country Coverage of Multicountry Patent Databases**[a]

Country, (treaty)	ISO code	Derwent WPI[b]	Caplus[c]	INPADOC[d]	FAMPAT[e]	ESP@CENET[f]	Other Databases
ALBANIA (EP * 1996) (PCT 1995)	AL						
ALGERIA (PCT 2000)	DZ						
ANTIGUA AND BARBUDA (PCT 2000)	AG						
ARGENTINA[g]	AR	1974–1976	1959, 1961, 1964	1973–	1973–	1973–1991	LATIPAT 1959–1995
ARIPO: INDUSTRIAL PROPERTY ORG. FOR AFRICA (PCT)	AP		2000–	1984–	1984–	1985–	
ARMENIA (EA 1996) (PCT 1991)	AM						
AUSTRALIA (PCT 1980)	AU	1963–1969 1982–	1927–	1973–	1966–	1945–	SURFIP
AUSTRIA (EP 1979) (PCT 1979)	AT [OE]	1975–	1907–	1973–	1969–	1825–	
AZERBAIJAN (EA 1995) (PCT 1995)	AZ						
BAHRAIN (GCC 1998)	BH						
BARBADOS (PCT 1985)	BB [BD]						
BELARUS (EA 1995) (PCT 1991)	BY						
BELGIUM (EP 1977) (PCT 1981)	BE	1963–	1928–	1973–	1964–	1923–	CLAIMS: 1950–1979
BELIZE (PCT 2000)	BZ						
BENGLADESH†	BD						
BENIN (OA 1983) (PCT 1987)	BJ [DA]						
BOLIVIA[g]	BO						LATIPAT 1989–96
BOSNIA & HERZIGOVINA (EP 2004) (PCT 1996)	BA			1998–	1998–	1998–	
BOTSWANA (AP 1994) (PCT 2003)	BW						

BRAZIL (PCT 1978)	BR	1975–	1957–1958, 1961, 1964, 1971, 1976–	1973–	1973–	1974–	
BULGARIA (EP 2002) (PCT 1984)	BG		2000–	1973–	1973–	1973–	
BURKINA FASO (OA 1983) (PCT 1989)	BF [HV,U-V]						
CAMEROON (OA 1982) (PCT 1978)	CM [KA]						
CANADA (PCT 1990)	CA	1963–	1910–	1973–	1973–	1940–	SURFIP 1920–
CENTRAL AFRICAN REPUBLIC (OA 1982) (PCT 1978)	CF [ZF]						
CHAD (TCHAD) (OA 1988) (PCT 1978)	TD [TS]						
CHILE[g]	CL		1919				LATIPAT 1959–1995
CHINA (PCT 1994)	CN	1985–	1985–	1985–	1986–	1985–	SURFIP 1985–
COLOMBIA (PCT 2001)	CO	LATIPAT 1968–1995					
COMOROS (PCT 2005)	KM						
CONGO (OA 1982) (PCT 1978)	CG [GF]						
COSTA RICA (PCT 1999)	CR	LATIPAT 1972–1995					
CROATIA (EP * 2004) (PCT 1998)	HR		2000–	1994–	1994–	1994–	
CUBA (PCT 1996)	CU		1974–	1974–	1974–1995		LATIPAT 1968–1995
CYPRUS (EP 1997) (PCT 1998)	CY		1975–	1975–	1921–1999		
CZECH REPUBLIC (EP 2002) (PCT 1993)	CZ	1993–	1993–	1993–	1993–	1993–	
CZECHOSLOVAKIA (PCT 1991)	CS	1975–1993	1955–1992	1973–1993	1973–1993	1973–1993	
DENMARK (EP 1990) (PCT 1978)	DK	1974–	1909–	1973–	1968–	1901–	

DJIBOUTI † (OA 1995)	DJ						
DOMINICA (PCT 1999)	DM						
DOMINICAN REPUBLIC[g]	DO						LATIPAT 1979–1995
ECUADOR (PCT 2001)	EC						LATIPAT 1986–1995
EGYPT (PCT 2003)	EG [ET]		2000–	1976–	1976–	1976–	
EL SALVADOR[g]	SV						LATIPAT 1968–1995
EQUATORIAL GUINEA (OA 2000) (PCT 2001)	GQ						
ESTONIA (EP 2002) (PCT 1994)	EE		2000–	2000–	1997–	1995–	
EURASIAN PATENT OFFICE (PCT)	EA			1996–	1997–	1996–	
EUROPEAN PATENT OFFICE (PCT)	EP	1978–	1979–	1978–	1978–	1978–	DELPHION: 1978– MICROPATENT: 1978–PHARM-SEARCH, 1983– SURFIP
FINLAND (EP 1996) (PCT 1980)	FI [SF]	1974–	1960–	1973–	1968–	1933–	
FRANCE (EP 1977) (PCT 1978)	FR	1963–	1900–	1973–	1902–	1903–	CLAIMS: 1950–1979 PHARMSEARCH, 1961–1973, 1983– MICROPATENT: 1981–
GABON (OA 1982) (PCT 1978)	GA						
GAMBIA (AP 1978) (PCT 1997)	GM						
GEORGIA (PCT 1991)	GE						
GERMANY, EAST	DD [DL] (EG)	1963–	1954–1994	1973–	1973–	1952–1999	
GERMANY, FEDERAL REPUBLIC (EP 1977) (PCT 1978)	DE [DT] (DS,G-E)	1963–	1899–	1973–	1877–	1879–	CLAIMS: 1950–1979 PHARMSEARCH, 1992–DELPHION 1968–MICROPA-TENT: 1989–

GHANA (AP 1978) (PCT 1997)	GH						
GREECE (EP 1986) (PCT 1990)	GR		2000–	1977–	1977–	1977–	
GRENADA (PCT 1998)	GD						
GUATEMALA[g]	GT						LATIPAT 1968–1995
GUINEA (OA 1990) (PCT 1991)	GN						
GUINEA–BISSAU (OA 1998) (PCT 1997)	GW						
GULF COOPERATION COUNCIL (GCC)[h]	GC			2002–	2002–		
HONDURAS†	HN						LATIPAT 1986–1995
HONG KONG (PCT 1997)	HK	2000–	1976–	1976–	1976–		
HUNGARY (EP 2003) (PCT 1980)	HU	1975–	1925, 1927–1931, 1933–1944, 1947–1949, 1957–	1973–	1994–	1973–	
ICELAND (EP 2004) (PCT 1995)	IS						
INDIA (PCT 1998)	IN	2004–	1935, 1941, 1941, 1943, 1946–	1975–	1975–	1975–1997	
INDONESIA (PCT 1997)	ID						
IRELAND (EP 1992) (PCT 1992)	IE [EI]	1963–1969 1995–		1973–	1973–	1930–	
ISRAEL (PCT 1996)	IL	1975–	1956–	1973–	1968–	1968–	
ITALY (EP 1978) (PCT 1985)	IT	1966–1969 1977–	1913–1914, 1919, 1925–1927, 1945–1972, 1999–	1973–	1973–	1933–	

IVORY COAST (OA 1982) (PCT 1991)	CI						
JAPAN (PCT 1978)	JP (J#) [JA]	1963–	1901, 1916–	1973–	1973–	1934–	DELPHION: 1976– SURFIP MICROPATENT: 1976–
KAZAKHSTAN (EA 1995) (PCT 1991)	KZ			1975–	1975–		
KENYA (AP 1994) (PCT 1994)	KE			1975–	1975–1989	1975–1989	
KOREA, NORTH (PCT 1980)	KP[KN]						
KOREA, SOUTH (PCT 1984)	KR [KS]	1986–	1995–	1978–	1978–	1978–	SURFIP 1979–
KUWAIT (GCC 1998)	KW						
KYRGYZSTAN (EA 1996) (PCT 1991)	KG						
LATVIA (EP*1995, 2005) (PCT 1993)	LV		1994–	1994–	1994–	1993–	
LESTHO (AP 1994) (PCT 1995)	LS						
LIBERIA (PCT 1994)	LR						
LIBYA (PCT 2005)	LY						
LIECHTENSTEIN (EP 1980) (PCT 1980)	LI						
LITHUANIA (EP *1994, 2004) (PCT 1994)	LT		1994–	1994–	1994–	1992–1994	
LUXEMBOURG (EP 1977) (PCT 1978)	LU	1984–	2001–	1973–	1946–	1945–	
MACEDONIA (EP * 1997) (PCT 1995)	MK						
MADAGASCAR (PCT 1978)	MG [MD]						
MALAWI (AP 1978) (PCT 1978)	MW			1973–	1973–	1973–1994	
MALAYSIA (PCT 1989)	MY			1953–	1971–	1953–1996	
MALI (OA 1984) (PCT 1984)	ML [MJ]						
MALTA[g]	MT			1967–	1968–	1967–1992	
MAURITANIA (OA 1982) (PCT 1983)	MR [MT]						

MEXICO (PCT 1995)	MX	1997–	1956, 1958, 2000–	1981–	1981–1995	1980–1995	LATIPAT 1968–2001
MOLDOVA (EA 1996) (PCT 1991)	MD		2000–	1994–	1994–	1994–	
MONACO (EP 1979) (PCT 1979)	MC		2000–	1975–	1975–	1957–	
MONGOLIA (PCT 1991)	MN [MO]			1972–	1972–	1972–1989	
MOROCCO (PCT 1999)	MA			1977–			
MOZAMBIQUE (AP 1978) (PCT 2000)	MZ						
NAMIBIA (AP 2004) (PCT 2004)	NA						
NETHERLANDS (EP 1977) (PCT 1979)	NL	1963–	1912–	1973–	1912–	1914–	CLAIMS: 1950–1979
NEW ZEALAND (PCT 1992)	NZ	1992–	2000–	1978–	1979–	1978–	
NICARAGUA (PCT 2003)	NI						LATIPAT 1979–1997
NIGER (OA 1982) (PCT 1993)	NE						
NIGERIA (PCT 2005)	NG						
NORWAY (PCT 1980)	NO	1974–	1907, 1909–	1973–	1968–	1936–	
OAPI: AFRICAN INTELLECTUAL PROPERTY ORG'N	OA	1993–	1966–	1960			
OMAN (GCC)(PCT 2001)	OM						
PANAMA†	PA						LATIPAT 1983–1995
PAPUA NEW GUINEA (PCT 2003)	PG						
PARAGUAY[g]	PY						LATIPAT 1968–1995
PATENT COOPERATION TREATY (PCT)	WO (WP)	1978–	1978–	1978–	1978–	1978–	DELPHION: 1978– MICROPATENT: 1978–PHARMSEARCH, 1989–
PERU[g]	PE						LATIPAT 1959–1995
PHILIPPINES (PCT 2001)	PH [RP]	1992–		1975–	1975–1994	1975–1997	
POLAND (EP 2004) (PCT 1990)	PL [PO]		1957–	1973–	1973–	1962–	

PORTUGAL (EP 1992) (PCT 1992)	PT	1974–	2000–	1976–	1976–	1976–	
QATAR[g] (GCC 1998)	QA						
ROMANIA (EP * 1996, 2003) (PCT 1979)	RO [RU]	1975–	1962–1963, 1966–	1973–	1973–1995	1973–	
RUSSIA (EA 1995) (PCT 1991)	RU	1993–	1908, 1911–1916, 1918, 1920–1941, 1943, 1993–	1993–	1993–	1993–	
SAINT LUCIA (PCT 1996)	LC						
SAINT VINCENT AND THE GRENADINES (PCT 2002)	VC						
SAN MARINO (PCT 2004)	SM						
SAUDI ARABIA† (GCC 1998)	SA						
SENEGAL (OA 1982) (PCT 1978)	SN						
SERBIA & MONTENEGRO (EP * 2004) (PCT 1997)	YU						
SEYCHELLES (PCT 2002)	SC						
SIERRA LEONE (AP 1999) (PCT 1997)	SL						
SINGAPORE (PCT 1995)	SG	1995–	2000–	1983–	1983–	1983–	SURFIP
SLOVAKIA (EP 2002) (PCT 1993)	SK	1993–	1994–	1993–	1993–	1993–	
SLOVENIA (EP *1994, 2002) (PCT 1994)	SI		2000–	1992–	1992–	1992–	
SOUTH AFRICA (PCT 1999)	ZA (SA)	1963–	1939, 1960, 1964–	1973–	1971–	1971–	
SPAIN (EP 1986) (PCT 1989)	ES	1983–	1946–	1968–	1968–	1968–	
SRI LANKA (PCT 1982)	LK [CL]						
SUDAN (AP 1978) (PCT 1984)	SD						
SWAZILAND (AP 1987) (PCT 1994)	SZ						

SWEDEN (EP 1978) (PCT 1978)	SE [SW]	1974–	1908–	1973–	1968–	1905–	
SWITZERLAND (EP 1977) (PCT 1978)	CH (SW)	1963–	1910–	1973–	1969–	1888–	
SYRIAN ARAB REPUBLIC (PCT 2003)	SY						
TAIWAN, REPUBLIC OF CHINA[g]	TW	1993–	1958, 2000–	2000–	2000–	1983–	SURFIP
TAJIKISTAN (EA 1995) (PCT 1991)	TJ			1998–	1998–	1998–	
TANZANIA (AP 1999)(PCT 1999)	TZ						
THAILAND[g]	TH						
TOGO (OA 1982) (PCT 1978)	TG [TO]						
TRINIDAD AND TOBAGO (PCT 1994)	TT				1994	1994–95	
TUNISIA (PCT 2001)	TN						
TURKEY (EP 2000) (PCT 1996)	TR		2000–	1973–	1973–	1973–	
TURKMENISTAN (EA 1995) (PCT 1991)	TM					1998–	
UGANDA (AP 1978) (PCT 1995)	UG						
UKRAINE (PCT 1991)	UA			2003–			
UNITED ARAB EMIRATES (GCC 1998) (PCT 1999)	AE						
UNITED KINGDOM (EP 1977) (PCT 1978)	GB (BR)	1963–	1901–1902, 1904–	1973	1909–	1863–	CLAIMS: 1950–1979 PHARMSEARCH, 1992–SURFIP MICROPATENT: 1916–
UNITED STATES (PCT 1978)	US	1963–	1905–	1973–	1920–	1859–	DELPHION: 1978– CLAIMS: 1950– PHARMSEARCH, 1983–SURFIP MICROPATENT: 1836–

URUGUAY[g]	UY						LATIPAT 1959–1995
USSR (PCT 1978)	SU (RU)	1963–1994	1940, 1943, 1945–1947, 1949–1950, 1952, 1955–1999	1972–1997	1978–1997	1940–1999	
UZEBEKISTAN (PCT 1991)	UZ						
VENEZUELA†	VE						LATIPAT 1959–1997
VIET–NAM (PCT 1993)	VN			1984–	1984–1997	1984–1997	
YUGOSLAVIA	YU			1973–	1973–	1973–1992	
ZAMBIA (AP 1978) (PCT 2001)	ZM [ZB]			1973–	1969–	1968–1995	
ZIMBABWE (AP 1980) (PCT 1997)	ZW [RH]			1980–	1980–		

[a]Country codes in brackets are obsolete ICIREPAT codes. Country codes in parentheses were used by Derwent Information Ltd. prior to the adoption of standard country codes. The obsolete codes remain on printed records, but have been replaced by current codes in most on-line databases. For Japanese patents, Derwent replaced the second character of the country code with the first digit of the publication year prior to 1992; this variant of the company code is still found in.online databases and printed publications.

Countries designated as members of the European Patent Organization, the Eurasian Patent Organization, the African Regional Industrial Property Organiz tion(AP), the African Intellectual Property Organization(OA) or the Patent Cooperation Treaty may be found as designated states in the records of EP, EA, AP, OA, AP, and WO documents, whether or not those countries are indexed separately in a database. The date of accession to the treaties varies from country to country, an refers to the year in which applications could be filed rather than the date of publication of the resulting patent documents. Extension of PCT applications to European Patent applications is possible for selected designated eastern European countries, most of which have become members of the European Patent Organization The year in which extension became possible is marked with an asterisk.

[b]For information about Derwent's starting dates for patent countries and kinds of documents, check their website at http://thomsonderwent.com/coverage/cpicovkinds/ for chemical patents, and http://thomsonderwent.com/coverage/epicovkinds/ for electrical patents.

[c]Details of Chemical Abstracts CAPLUS patent coverage can be found at http://www.cas.org/EO/patyear.html

[d]Details of INPADOC coverage are available from http://www.european–patent–office.org/inpadoc/statistics_dwld.htm

[e]Details of FamPat coverage may be found at http://www.questel.orbit.com/EN/customersupport/Userdoc/Fctsht/FamPat.pdf

[f]Details of ESPatCENET coverage may be found at http://ep.espacenet.com/ep/en/helpV3/coverageep.html

[g]Not a member of the Patent Cooperation Treaty as of August, 2005.

[h]The Gulf Cooperation Council was established by seven Middle Eastern countries for joint patenting, effective from 1998.

Table 4. **Scope of Derwent CPI and WPI**

		Number of basic patent references, $\times 10^3$			
CPI Section	Subject content	1973	1983	1993	2003
A	PLASDOC: polymers	28.3	39.7	63.1	83.0
B	FARMDOC: pharmaceuticals	6.8	10.1	18.9	56.3
C	AGDOC: agricultural chemicals	4.0	5.1	6.2	9.8
D	Food, biotechnology, detergents, cosmetics, etc	8.3	15.5	29.4	68.4
E	CHEMDOC: general chemicals	15.8	19.6	27.8	33.4
F	Textiles, paper, cellulose	11.6	10.1	15.1	13.3
G	Printing, coating, photographic chemistry	5.4	10.0	21.2	18.8
H	Petroleum	4.3	8.1	9.0	12.4
J	Chemical engineering	6.6	14.4	19.6	19.8
K	Nuclear, explosives, protection	2.2	3.4	4.3	3.6
L	Glass, refractories, ceramics, electrochemistry	9.0	26.6	47.8	60.6
M	Metallurgy	13.1	25.2	32.3	22.9
Total CPI		*81.0*	*128.0*	*183.5*	*226.2*
Total WPI		*81.0*	*302.4*	*413.5*	*902.7*