

Patent Searching: Back to the Future

How to Use Patent Classification Search Tools to Create Better Searches

Michael White
Librarian for Research Services
Queen's University
michael.white@queensu.ca

Abstract

The proliferation of free patent databases on the web over the last fifteen years has dramatically improved access to patent information. Ironically, as access to patent information has improved, emphasis on teaching engineering students how to search the patent literature using patent classification has waned. Recently published engineering textbooks and articles recommend that students search patent databases using keywords. However, keyword searching is unreliable and produces incomplete results. Engineering students who learn to search patent databases using keywords risk developing poor information management skills. This paper will illustrate the problems that arise in keyword searches and demonstrate strategies for teaching students how to use patent classification to improve their search results, thus encouraging the development of good information retrieval and management skills.

1 Introduction

[The engineer] must know about patents and know something of patent law, else he is not qualified for full duty and is not fully caring for the interests entrusted to him.

William Macomber, *Engineers' Handbook on Patents*, 1913[1]

Knowledge of patents and intellectual property (IP) in general is essential for the modern professional engineer. Today's global economy is a highly competitive environment powered by technological innovation. In 2007 alone, more than 1.85 million patent applications were filed worldwide. [2] Engineers must know how to protect the IP that they create, whether they work for a small start-up company or large multinational firm. In addition, engineers must be aware of and respect the IP rights of others or else risk costly infringement mistakes.

Most importantly, engineers should know how to tap into the rich technical information contained in the vast and rapidly growing body of patent literature. [3]

The value of patents as sources of technical and scientific information cannot be overstated. Patents detail innovations in every field of technology during the last 200 years. Patents disclose information that is not published in scientific journals or conference proceedings. Studies have shown that up to 80 percent of the information contained in patents is not published elsewhere. [4][5] Increasingly, patents are a window onto research and innovation in developing countries. China is a burgeoning hub of patent activity, due, in part, to patent law reform and government policies encouraging domestic research and innovation. [6] In 2009, China and the Republic of Korea accounted for 10.2 percent of new international patent applications, up from 2.5 percent in 2000. [7]

Numerous articles published since the 1980s describe methods of teaching engineering students about patents and patent searching. [8][9][10][11][12] Some address the use of patent classification to teach students effective search strategies. However, many only briefly mention classification or recommend keyword searches. [2] This is unfortunate because a basic understanding of patent classification will greatly improve search results.

2 Problems with keyword searching

Having grown up using internet search engines, engineering students are thoroughly familiar with keyword searching. However, keyword searching in patent databases is highly problematic. [14]

First, public patent databases generally contain full-text patent data from the 1970s forward. Titles and abstracts may be available for earlier patents or not at all. For example, U.S. patents prior to 1976 can only be retrieved in the USPTO patent database by number, date and classification. Almost four million

U.S. patent documents, nearly half the total number issued, are invisible to keyword searches.

The vast majority of patents are improvements on existing technologies. Even seemingly futuristic inventions can have surprising historical precedents. For example, the original patent for the TASER® electronic stun gun, which was issued in 1974, cites a patent for an electric harpoon that was issued in 1852! [15] Among 350 cited patents in 25 recently issued patents related to wind turbines examined by the author, 21 percent were issued prior to 1976 and 14 percent were issued before 1950. The oldest cited patent was issued in 1880.

Another problem with keyword searching is the growing number of patent documents published in foreign languages. Multinational patent databases such as esp@cenet and PatentScope cover patents in many languages, including Chinese, Japanese and Korean. Although some patent titles and abstracts are eventually translated into English, many are not. Patent titles are infamous for their use of broad terms such as “device” or “method” and synonyms. For example, a patent for a mousetrap might be titled “Rodent Extermination Device” and a patent for a toy ball a “Generally Spherical Object for Amusement”. In such an environment keyword searching is next to useless.

3 Patent classification

The organization, storage and retrieval of patent documents have been significant concerns of patent offices for almost 200 years. One of the basic tenets of patent law is that patents are granted only for new and unobvious inventions. Novelty can only be determined by a thorough search of the prior art, which includes all previously issued patents and published non-patent literature. Patent offices developed classification systems in the 19th century in order to cope with the growing volume of patents and non-patent literature. The three most commonly used patent classifications today are the U.S. Patent Classification, International Patent Classification and European Classification. It should also be noted that these classification systems are non-proprietary and available to the public. The list of resources at the end of this paper gives the website address for each of these systems.

3.1 U.S. Patent Classification (USPC)

The modern USPC emerged around 1900 when the U.S. Patent Office published the first formal rules of classification. In the beginning, the USPC system consisted of about 235 classes; the number of

documents classified under it was about 700,000 U.S. patents and 1.25 million foreign patents. Today, the USPC consists of 450 classes, including 35 design classes and one class for plant patents, and approximately 150,000 subclasses. [15] More than 8 million U.S. patent documents are assigned 28 million USPC classifications, an average of 3.4 per document. [16]

The USPC is a horizontal classification system. Classes are arranged in numerical order from Class 2, Apparel through Class 976, Nuclear Technology. Design classes are numbered D1 through D34, plus D99 for miscellaneous designs; the class for plant patents is designated PLT. The USPC is supported by an alphabetical index of common terms and corresponding USPC classifications; each subclass has a definition that provides greater detail about the scope of the subject matter classified and useful “see also” directions for searching other related subclasses. At the class level there is no relationship between proximity and subject matter. It is not uncommon for some technologies to span two or more class numbers. For example, surgical instruments are classified in classes 128, 600, 601 and 602. Individual class schedules are organized in a hierarchical fashion. USPC codes must consist of a class and subclass number. A typical USPC hierarchy is shown below.

Table 1. USPC hierarchy for 429/31

Level	Number	Title
Class	429	Chemistry: electrical current producing apparatus, product, and process
Subclass (main line)	12	Fuel cell, subcombination thereof or methods of operating
Subclass	. 30	Solid electrolyte
Subclass	. . 31	Tubular

In example above, 429/31 encompasses fuel cells comprising solid electrolytes arranged in a tubular configuration.

The USPC is continuously revised in order to accommodate new technologies and increasing complexity in established technologies. These changes are published in official bulletins known as Classification Orders which are available on the USPTO website. The USPTO updates the classification data in its internal and public databases bimonthly.

3.2 International Patent Classification (IPC)

The International Patent Classification (IPC) is based on an international treaty that was concluded in Strasbourg, France in 1971 and entered into force in 1975. As of July 2008, 58 states were party to the Strasbourg Agreement, which is administered by the World Intellectual Property Organization (WIPO). The IPC is used by more than 100 national patent offices, four regional offices and the WIPO. It is estimated that IPC codes have been assigned to more than 50 million patent documents since 1975. Prior to 2006, new editions of the IPC were released every five years. The IPC is now revised on an ongoing basis with new editions published at least once a year.

The IPC is a hierarchical classification system. The top level consists of eight sections, A through H, which are divided into 70,000 subdivisions called classes, subclasses and groups. [17]

3.3 European Classification (ECLA)

ECLA is an enhanced version of the IPC developed by the European Patent Office (EPO) to organize European and foreign patent documents and scientific literature. ECLA uses the same hierarchical structure as the IPC but has almost twice (134,000) the number of subdivisions, allowing for greater granularity in searching. ECLA codes are not printed on European patent documents but they can be searched in esp@cenet. Approximately 30 million patent documents and 1 million NPL references are classified under ECLA. [18]

Table 2. IPC and ECLA hierarchy for H01M 8/10

Level	Number	Title
Section	H	Electricity
Class	H01	Basic electrical elements
Subclass	H01M	Process or means for the direct conversion of chemical energy into electrical energy
Main group	H01M 8/00	Fuel cells
Subgroup	H01M 8/10	Fuel cells with solid electrolytes
ECLA subgroup	H01M 8/10B2U	Undulated, corrugated, curved or waved-shape membrane electrode assemblies (MEA)

3.4 National classification systems

A number of patent offices developed their own patent classification systems. In recent years these national classifications largely have been abandoned in favour of the IPC. For example, the Canadian Intellectual Property Office discontinued classifying patents under the Canadian Patent Classification (CPC) in 1989. Japan uses two locally developed classification systems, FI (File Index), which is based on the IPC and contains 170,000 subdivisions, and File Forming Terms (F-Terms), an indexing system based on 2,800 themes and 350,000 terms. National classifications can be useful for searching historical patent documents that are not assigned IPC codes. For example, in the Canadian Patents Database, IPC codes are not searchable prior to 1972. However, pre-1972 Canadian patents can be retrieved using CPC codes. F-Terms can be used to retrieve Japanese patents back to 1885.

4 Classification in patent databases

Selecting a patent database is the first and most important step in the search process. Not all patent databases, as we shall see, are created equal. Selecting a patent database based only on its superficial qualities, such as popularity and ease of use, and with little or no consideration of its scope, organization and currency will result in inaccurate and incomplete searches.

The USPTO launched the first large public patent database on the internet in November 1995. It contained bibliographic records for approximately 1.6 million U.S. patent documents dating back to January 1, 1976 and was updated weekly, on Tuesdays, when new patents issued. Within a few years the USPTO expanded the database to include the full-text and/or images of all patents issued back to 1790. Today, the USPTO's databases contain more than 10 million patent documents, with approximately 500,000 new documents added each year.

The late 1990s saw a proliferation of web-based patent databases spurred in part by the availability of increasingly powerful (and inexpensive) computing technology and the relatively low cost of obtaining raw patent data. By the end of the decade, most of the major patent offices had launched their own public patent databases. The Canadian Intellectual Property Office launched its Canadian Patents Database in October 1998. These official patent databases were soon joined by dozens more created by entrepreneurs, enthusiasts, collectors, academic researchers and librarians.

Today there are more than 100 free and subscription patent and patent-related databases on the web ranging in size and scope from small subject-focused collections, such as Georgetown University's database of DNA patents, to huge, multinational patent databases. Perhaps the largest public database is the European Patent Office's esp@cenet system, which contains more than 60 million patent documents from 80 countries, including the U.S. and Canada, and one million references to non-patent literature. Other popular patent databases include Google Patents, Patent Lens and FreePatentsOnline. Patent classification coverage varies by database.

Table 3. National and IPC classification coverage in public and free patent databases

Database	Producer	National	IPC
U.S. PatFT/AppFT	USPTO	1790-present	~1900-present
Canadian Patents Database	CIPO	1869-1990	~1972-present
Esp@cenet (80+ countries)	EPO	~1920-present	~1920-present
IP Digital Library (JP)	JPO	1885-present	1885-present
PatentScope (WO)	WIPO	NA	1978-present
FreePatentsOnline (EP, JP, US, WO)	FPO.com	1975-present (USPC)	1978-present
Patent Lens (AU, EP, US, WO)	Cambria	NA	NA
Google Patents (US)	Google	?	?

Note: AU, Australia; EP, European; JP, Japan; WO, PCT application.

As this table shows, patent databases produced by patent offices generally have better classification coverage than free patent databases. Searchers should also be aware that classification data in free patent databases may not be up to date. For example, Google Patents does not provide current classification data although it claims to do so. Patent Lens does not support classification searching at all. Instead, the developers of Patent Lens have chosen to focus on search tools most useful to researchers in the life sciences, such as the ability to search DNA and protein sequences in patent applications. Some free patent database search engines have a difficult time interpreting the special formats of some classification codes.

Several patent databases have integrated classification tools that greatly enhance the efficiency and effectiveness of searches. One of the best is esp@cenet, the multinational patent database produced by the European Patent Office. The next section will demonstrate how to use esp@cenet's classification search tools to improve patent search results.

5 Patent searching: back to the future

Prior to the availability of patent databases on the web, searching patents was a time-consuming and physically demanding affair. Searchers were required to conduct searches at a patent office or one of the few public or university libraries that had collections of patent documents. Print search tools consisted of complicated patent classification manuals, indexes and gazettes, which were often weeks or even months out of date. Commercial online patent databases appeared in the early 1980s, followed by CD-ROMs a few years later. However, these electronic search tools were available only at a few select libraries and, in the case of online patent databases, very expensive and difficult for inexperienced searchers to use.

Modern web-based patent databases have overcome or reduced many of these obstacles. Anyone with access to the web and a basic knowledge of patents and database searching can learn to use patent classification.

The standard patent search process is given in Fig. 1 below.

Fig. 1 Standard patent search outline modified for use in esp@cenet

Step 1: Write a brief but accurate description of the invention or technology, noting important keywords and synonyms.

Step 2: Go to the Advanced Search form. Search keywords and synonyms in the Title and Abstract field. Sort search results by ECLA classification. Examine retrieved documents and note classification codes on patents closely related to the invention.

Step 3: Click on the selected ECLA classification in order to determine its scope. If appropriate, copy into the search form.

Step 4: Search selected ECLA classification and examine retrieved patent documents. If documents are not appropriate, return to Step 2 and select a new classification.

Step 5: Supplement ECLA search by searching appropriate IPC codes. This will retrieve relevant documents that are not assigned ECLA codes.

5.1 Search 1: Collapsible bicycle helmet

Modern bicycle helmets are effective but bulky and difficult to store. A collapsible bicycle helmet would be much easier to carry in a backpack or store in a desk drawer.

The logical starting point for this example would be a keyword search on the words “collapsible” and “helmet”. Truncating both words (“collaps* AND helmet*”) will increase the number of documents retrieved. This search retrieves 81 patent documents, many of which are not relevant, e.g. “Backpack for Snow Skis and Boots”.

Sorting the search results by ECLA code makes it easier to identify relevant patents. An examination of the sorted hit list reveals several good candidates assigned ECLA code A42B 3/32B. Clicking on the classification reveals the ECLA hierarchy; A42B 3/32B is the code for collapsible helmets. Searching this code retrieves 133 patents, including several in languages that would not have been retrieved by a keyword search in English.

Since not all patent documents in esp@cenet are assigned ECLA codes, it is a good idea to supplement the ECLA search with a search of the corresponding IPC code for collapsible helmets and helmets made of separable parts, A42B 3/32. This search retrieves more than one thousand patents because it is slightly broader in scope. The number can be reduced by limiting the search by year range.

5.2 Search 2: Ocean wave electricity generator

The oceans of the world are a potential source of clean, renewable energy in the form of electricity generated by the motion of waves acting on floating power plants. Currently, there are pilot “wave energy” farms in operation off the coasts of Portugal and Scotland. Research continues on more efficient technologies for converting wave energy into electricity, many of which have been disclosed in patent documents. This search is more complicated than the previous example because the descriptive keywords and their synonyms, e.g. “wave”, “electricity”, “power”, “energy”, “generator”, etc. have many different applications and meanings. For example, a keyword search in the title or abstract for “wave” AND “electricity” AND (“generat* OR “harvest*” OR “captur*”) retrieves more than 1,300

patent documents, many of which are not related to the ocean wave electricity generators.

This problem can be overcome using esp@cenet’s integrated Classification Search. The Classification Search maps any keyword search to the most appropriate classifications in the ECLA hierarchy. Candidate codes are presented as a ranked list. For example, searching “wave” AND “electricity” and (generat* OR harvest*)” retrieves F03B13 as the top candidate. This code is found under F03B, Machines or Engines for Liquids, and includes “Adaptations of machines for special use.” Clicking on the title or code will reveal the entire ECLA hierarchy, which can be browsed in order to locate the most appropriate subdivision. In this case, the most appropriate classification is F03B 13/14, machines drive by or using ocean wave energy.

Below F03B 13/14 are a number of more specific subdivisions. For example, F03B 13/14B includes wave energy machines combined with static energy collectors. This example shows that by browsing the ECLA hierarchy is it possible to identify classification for very specific features or designs.

6 Conclusions

Patent databases are an extremely valuable (and free) source of information on technology and engineering design. Keyword searching in patent databases is problematic and should be avoided when a comprehensive search is required. Patent classification systems are powerful tools for organizing and searching the patent literature. Novice searchers can greatly improve their search results using integrated classification search tools available in public patent databases.

7 List of Patent Databases and Classification Search Tools

CIPO Canadian Patents Database

<http://patents.ic.gc.ca/cipo/cpd/en/introduction.html>

DEPATISnet (German Patent and Trademark Office)

<http://depatismet.dpma.de/>

EPO esp@cenet

<http://ep.espacenet.com>

EPO ECLA Classification

http://v3.espacenet.com/eclsrch?classification=ecla&locale=en_EP

FreePatentsOnline

<http://www.freepatentsonline.com/>

Google Patents

<http://www.google.com/patents>

Patent Lens

<http://www.patentlens.com/>

USPTO PatFT and AppFT

<http://www.uspto.gov/patft/index.html>

USPTO Classification Tools

<http://www.uspto.gov/go/classification>

WIPO PatentScope

<http://www.wipo.int/patentscope/en/>

WIPO International Patent Classification

<http://www.wipo.int/classifications/ipc/en/>

8 References

[1] W. Macomber, *Engineers' Handbook on Patents*, Little, Brown & Co., Boston, 1913.

[2] World Intellectual Property Organization, *World Intellectual Property Indicators 2009*, WIPO, Geneva, 2010.

[3] G. C. Andrews, *Canadian Professional Engineering and Geosciences: Practices and Ethics*, 4th ed., Nelson Education, Toronto, 2009.

[5] J. Oppenheim and C. Oppenheim, "The Overlap of U.S. and Canadian Patent Literature with Journal Literature", *World Patent Information*, 1979, pp. 77-80.

[6] P. J. Terragno, "Patents as Technical Literature", *IEEE Transactions on Professional Communication*, 1979, pp. 101-104.

[7] E. Y. Zhou and B. Stembridge, *Patented in China: the Present and Future State of Innovation in China*, Thomson Reuters, London, 2010.

[8] World Intellectual Property Organization, *PCT Yearly Review: the International Patent System in 2008*, WIPO, Geneva, 2008.

[9] O. J. Whittemore, "Patents: a Tool for Teaching Design," *Engineering Education*, 1981, pp. 299-301.

[10] R. Soetendorp, "Patent Information in the Academic Context", *World Patent Information*, 1996, pp. 219-226.

[11] C. A. Garris, "The United States Patent System: An Essential Role in Engineering Design Education", *Journal of Engineering Education*, 2001, pp. 239-246.

[12] J. Nazemetz, P. Rossler, M. High, and K. High, "Why Reinvent the Wheel? The U.S. Patent and Trademark Office as a Design Tool", *Annual ASEE Conference and Exposition*, ASEE, Washington, DC, 2007.

[13] M. Macmillan and L. Shaw, "Teaching Chemistry Students How to Use Patent Databases to Glean Patent Information", *Journal of Chemical Education*, 2008, pp. 997-999.

[14] S. Adams, "Patent Searching Without Words: Why Do It? How to do It?", *Free Print Newsletter*, 130, 2003, <http://www.freepint.com>.

[15] J. Cover, *Weapon for Immobilization and Capture*, U.S. pat. 3,803,463, 1974.

[16] I. J. Rotkin and K. J. Dood, *A History of Patent Classification in the U.S. Patent and Trademark Office*, Patent Documentation Society, Arlington, Virginia, 1999.

[17] T. Thomas, private e-mail, May 15, 2010.

[18] World Intellectual Property Organization, *International Patent Classification Guide*, WIPO, Geneva, 2009.

[19] European Patent Office, esp@cenet system online help, accessed May 15, 2010.