Patent Search Strategies: Keywords or Classifications?

Lately some new tools have become available for patent research, and some old ones have been improved. Meanwhile, debates still rage on which are the best techniques to use - the professional searchers using patent classification methods ridicule the numbskulls who want to search only by keywords.

The reality is that in order to conduct a serious search, all possible relevant techniques should be applied. In many situations, intelligent use of keywords combined with classifications is the principal way to go.

Consider the plight of a searcher looking for business method patents. The narrowest International Patent subclass on this subject is G06F-017/60, which on the Derwent World Patent Index database includes over 48,000 inventions. Only in the European Patent Office adaptation of the IPC are a dozen or so subclasses indicated to provide a further division of this bottomless swamp of patent documents. The US patent classification is useful, but most of the interesting prior art in this area is spinning in cyberspace as PCT applications, not as issued US patents.

In this and many other instances the patent classification system is years behind technology development, and many patent classifications are simply black holes bearing little signs like "Misc. Computer Gadgets". In such cases, the main approach is the "shotgun search", using one or more patent classifications as screens to employ a wide range of keyword search strategies. Such search strategies must make intelligent use of proximity operators to retrieve documents that include search terms meaningfully grouped together in relation to each other.

Often such searches still produce hundreds of hits for potentially relevant patents. Reviewing several hundred documents an hour is best done with a good computer connection to a "text only" version of a service such as Lexis Nexis, that allows rapid display of keywords in context of full text.

Why full text? It is generally appreciated that patent abstracts are superficial and badly done, and the real nuggets of truth may be buried under the boilerplate somewhere in the summary, claims or description of embodiments of a patent document. Often the

description of the prior art highlights some ancient patent beyond the reach of electronic indexing.

KEYWORD SEARCHING

The main pitfall of keyword searching is the English Language as it is used and abused. However, here in detail are the specific problems:

Poor Translations: Probably 40% of US patents are filed from abroad, by people who may not use your favorite set of buzzwords for describing their technology. A great many Japanese translations are execrable, on the level of blind Martians describing an elephant. Machine translations are not always the worst source of garbage.

Different Spin: One mans plant fungicide is another mans crop growth promoter. Patent agents are hired to make the humdrum sound novel.

Too Many Synonyms: In the realm of chemistry, a compound may have twenty names, depending on the country or field of use. Use Chemical Abstracts to search with the CA registry number.

Generic VS Specific: There is always a chance that a patent out there dominates because it claims whole genera of applications or substances. Who has not cringed at the dreaded words "Alkali metals or salts thereof" or "Group VII elements" or "Lanthanide complexes". Search for general terms as well as specific details.

No Spelling Standards: Some people assume all databases are cast in American English. The World Patent Index, one of the best international databases, is created in the UK and full of "tyres", "lorries", "lifts", "bitumen" and many other British terms.

Really Bad Titles and Abstracts: No regulation decrees that these fields should aid in retrieval, and not all databases are created equal - some vendors such as Derwent provide titles and abstracts enhanced by skilled indexers, whereas most others simply dump raw data from patent office files into their databases.

Innovative Lexicography: Everyone drafting patents is allowed to create novel terminology to describe inventions and ascribe new meanings to words. The system encourages bizarre descriptions.

Errors and Omissions: There are thousands of examples of obvious mistakes in titles and names of inventors and assignees as well as

other important data such as cited patents. Very few of the errors in issued patents are ever corrected by the database suppliers, who often add their own. Worldwide the rule seems to be GIGO - Garbage In…

Date Limitations: Most databases only go back about 30 years or so for searchable abstracts or text. Before that, there is only patent classification, unless you use such sources as Lexis Nexis, MicroPatent or Chem. Abstracts that have lately provided back-file coverage.

Survival Strategies: Use utilities such as Delphion's Text Clustering, which allows comparisons of keywords found in groups of patents in order to select best terms for searching. Otherwise bookmark online thesauri and technical dictionaries. Since technical terms tend to morph over time, it is a good idea to browse the index terms for a given subject area over the years (such as the INSPEC thesaurus for electrical engineering terminology or MESH for medical subjects) to review usage changes in the technical literature. Sadly few patent databases have rigidly controlled vocabularies including role qualifiers and weighted terms, such as APIPAT. Systematic keyword searching involves drafting a grid of narrow, broad, and related terms, grouped together in proximity and related to other groups of terms expressing functionality or application. The terms must then be searched in rotation including every conceivable permutation and combination.

CLASSIFICATION SEARCHING

The theory of classification searching is based on professionals in patent offices making judgment calls to classify patent documents according to the subject matter claimed. Thus, classified patent files should be a perfect search tool. Unfortunately, the decisions of classifiers are often subjective, incomplete and random. Some classification searching pitfalls are:

Wild Variations: The International Patent Classification gets a wide range of interpretations in different jurisdictions. Some folks seldom go beyond the "animal, mineral or vegetable" level of classification. Others bury you with irrelevancies.

Stagnant Classification: In quickly developing technologies the IPC Classification issued every 5 years may be unhelpful. The European Patent Office revisions of ECLA (European Classification) are timelier, and the USPTO makes a good effort to keep current. The USPTO has started including ECLA subclasses in the Manual of Classification.

Obsolete Classifications: In many free databases and even in some expensive ones, there is no linkage between obsolete classifications and recently changed ones. The USPTO is quite religious in providing this guidance. It is always important to toss a relevant old US patent number into the mill to discover the current classification.

Minimal Classification: In some databases such as that for Canadian patents on CIPO's website, the lack of classification is a serious hindrance.

Survival Strategies: Use all the free classification tools on the Internet if that is the only access you have. The USPTO offers a searchable Manual of Patent Classification and Index to the US Patent Classification linked to Classification Definitions. Espacenet now has a tool for keyword searching of the text of the excellent ECLA classification. Unfortunately, the arcane patent jargon used in the classification manuals will not help you find subjects such as "pencils" unless you are looking for "writing implements." If you use the commercial databases there are utilities for statistically ranking classifications within search results, or text searching the classification manuals. Always compare the classifications allocated to important patents, whether USPTO, ECLA, or IPC.

In ancient history (a couple of decades ago) most patent searching was done manually. Searchers flipped through classified collections of paper copies in the monastery-like search rooms of the major patent offices. When most inventions were mechanical and viewing drawings was the quickest way to review patents, this worked well. Now the US patent office has served notice that it will soon turf out the paper, and everyone will have to search US patents online.

Given the enormous volume and complexity of today's patent data, where 100 page patents are commonplace, a searcher must become a data-miner, and use all the best computer tools he or she can afford. Thoroughness and persistence is critical. A good searcher will use competitive intelligence tactics, checking newsbases, new product announcements, technical literature and press release archives for evidence or prior art.

Citation searching of EP, PCT and US documents is always very important. Technical literature searches often turn up valuable leads since academic inventors often file patents following their research papers. Searching for cited research articles or cited scientific authors in patent data is often fruitful in identifying key patents.

Many have pointed out that there can be no magic bullets or guarantees in patent searching, given that the patent literature is generally such an inconsistent and messy system. Compared to the academic network of refereed journals with exhaustive and meticulous citations documenting the evolution of every idea, patents are a bibliographic nightmare. As Greg Aharonian has mentioned once or twice, the massive lack of citation references for many US patents indicates a serious failure in the examination effort.

Stu Kaback emphasizes there is no single best way to search patents. He notes there are a lot of resources with different capabilities, and a good patent searcher will learn what those capabilities are and use them in combination judiciously.

For the average searcher in a small high technology operation whose searching budget is near the bottom of the price continuum, there are slow improvements in database access, coverage and quality. However, increasing amounts of money are needed to access basic effective searching tools and the necessary training and support.

Bibliography

Schwander P. An evaluation of patent searching resources: comparing the professional and free on-line databases. World Patent Information. 22(3), 147-165, 2000

Gerald R.Black. "Keyword Patent Searching Online - a Workbook" 120 pages www.keypatent.net