

Patents on the Internet versus Patents Online: A Snapshot in Time[†]

Nancy Lambert*

Chevron Business Products and Services, P.O. Box 1627, 100 Chevron Way, Richmond, California 94802

Received July 13, 1998

Internet-based patent information resources have grown considerably in the past few years. Some of them offer search and display capabilities not available through commercial online patent databases, whereas the indexed online databases and the hosts that mount them offer in-depth searching, particularly in chemical areas, and other capabilities still not available on Internet-based databases. This article gives a state-of-the-art summary of Internet and online resources currently available and looks at capabilities unique to both.

INTRODUCTION

I am aware that comparing Internet and online patent searching capabilities is comparing apples and oranges (if not indeed grapes and watermelons), because they are aimed at fundamentally different audiences. The Internet is the ultimate end-user tool, on which people will browse for hours, unconcerned about connect-hour charges (which do not normally exist). Search engines are simple, or at least self-teachable. On the other hand, online databases attract professional searchers, who usually formulate searches in detail beforehand and try to get on and off line as quickly as possible, partly because they are usually chalking up connect-hour charges, and partly because they must spend their own time efficiently—they probably charge it back to a search client. Online search engines are powerful, but searchers need training, sometimes extensive training.

Nevertheless, some comparisons of Internet versus online capabilities are in order because people in decision-making positions sometimes look at the Internet as the repository of all knowledge and wonder why the more expensive online databases are still necessary. The question, as searchers well know, is not whether the patents are out there on the Internet. The question is whether you can extract exactly the patents that you want from the millions of patents out there. And because for any complex subject search the answer to that question is a resounding “No,” the next question is, how much closer to that goal can you come with online resources?

In this article I will not address the gray area occupied by systems that provide more or less user-friendly Internet gateways to access online databases, such as Dialog Select, Dialog Web, and STN Easy. I will address six Internet-mounted databases and two others that are not exactly Internet mounted but come close. The Internet databases are the Community of Science, a relatively new site; the IBM patent server; Derwent's Patent Explorer; MicroPatent's Patent Web; QPAT-US; and the USPTO (United States Patent and Trademark Office) front-page database. The others are Chemical Abstracts Service's Chemical Patents Plus, which is actually a gateway to the USPatFull database on

STN; and the Corporate Intelligence database, which is mounted on their own computer. For brevity, I will refer to all of these as Internet sites in comparing them with online databases.

CONTENT

As to the first question just posed, are the patents indeed out there on the Internet? Only partly. Older U.S. patents and most international patents are not yet available in Internet-mounted databases. In contrast, many online databases provide in-depth searching of both U.S. and international patents going back as far as the early 1960s (to 1950 for U.S. chemical patents). So do some specialized databases in, for instance, petroleum and pharmaceutical areas.

Most of the Internet databases include only what they get from the U.S. Patent and Trademark Office: U.S. patents back to 1974, partially back to 1971—what the USPTO has made available in machine-readable form. Only three resources go beyond this as of now. MicroPatent has mounted the EPO (European Patent Office) and PCT (Patent Cooperation Treaty) patent gazettes back to 1978, and Patent Abstracts of Japan (JAP10) back to 1994, on its web site with bibliographic information and English language title and abstract text, optionally searchable together with U.S. Official Gazette back to 1964. Patent Explorer includes some EPO patents full-text and others with bibliographic information and abstracts. Finally, Corporate Intelligence has produced machine-readable text for U.S. patents back to 1945.

On the international scene, the EPO will soon mount a database of its own patents. Not much information is available about this yet, but apparently the database is aimed at end-users and will have very limited search capabilities. The EPO has stated that it does not want to compete with the commercial databases. The EPO will sponsor inclusion of the recent 2 years' data from each member country, and individual countries may optionally add further back data at their own expense. The service will provide some links to patent full-text.

Starting early in 1998, the PCT Gazette is searchable (and viewable, with abstracts and front-page drawings) from Jan 1, 1998 onward on its own Internet site. Older gazettes will be added. Also, some individual country patent offices are making their patents searchable on the Internet. For

* Corresponding author. Telephone: 510-242-3161. Fax: 510-242-5621. E-mail: nela@chevron.com.

[†] Paper presented at the ACS Meeting CINF Symposium on Patent Information, Dallas, Texas, March 30, 1998.

instance, Patent Abstracts of Japan (laid-open patent applications from Japanese applicants only) is now available free on the Internet and searchable 1976–date by English language title/abstract text, International Patent Classes, and bibliographic information. Canadian patents, both laid-open and granted, are searchable (biblio, IPCs, and title text only) from 1989 to date via the Canadian Patent Office web page. More countries will probably be available by the time you read this. Keep in mind that when you search these web sites, you are searching one country's patents at a time, with no way of matching international patent family members. Compare this with the multi-country coverage of, for instance, the Derwent World Patents Index.

TEXT SEARCHABLE

Both Internet and online databases vary as to how much text they make searchable. Most offer title and abstract text. On the Internet, the IBM site also offers full claims text. Chemical Patents Plus, Patent Explorer, and QPAT-US offer full-text searching. Online, the CLAIMS and USP databases make full claims of U.S. patents searchable, and both Dialog and STN offer their own full-text U.S. and EPO patent databases (STN's EPO file only 1996-date).

SUBJECT ACCESS

Internet databases for the most part provide free-text searching of varying quantities of patent text. But, as anyone knows who has read through a few patents, the text can make for very poor searches. Patent writers are their own lexicographers, and they have considerable leeway in what they may name chemicals and other patentable entities. Patent text is inconsistent; it is sometimes deliberately vague; it can describe parts of an invention in generic terms or in terms of their components, without ever actually naming them; and it often reflects the culture of the patentees. A free-text search, even of full text, virtually never provides a complete search of a subject.

Probably the most important feature missing from the Internet patent databases is the indexing on which most professional searchers depend, especially for chemical patents. The exceptions are Chemical Patents Plus, which offers both CAS indexing (for chemical patents only) and some patent classification catchword text; and Patent Explorer, which includes broad Derwent subject classes.

Bottom line, no one can do a comprehensive search for chemical patents using patent free text. You must have access to the various indexing systems that translate text into concepts and chemicals into searchable structures. For my own company's interests, I make use of the petroleum subject indexing of both the Petroleum Abstracts and American Petroleum Index (API) databases, as well as the chemical and polymer indexing in the API databases, the Derwent World Patents Index (DWPI), the IFI Comprehensive Index, and of course Chemical Abstracts. My colleagues in other fields, such as pharmaceutical or photographic chemistry, can add other specialized databases to that list.

PATENT CLASSIFICATIONS

Then there are patent classifications, both U.S. and International.

International Patent Classification. Standard IPCs are available in various databases, both Internet and online, and present the same problems wherever you search them. Class hierarchies are revised in batch about once every 5 years, old patents are not reclassified, and you must check old editions of the IPCs if you plan to search them over an extended time period.

European Patent Classification. ECLA (European Classifications), the European Patent Office version of the IPCs, are more detailed than the IPCs. They are revised in a continuous process, and old patents are reclassified in new classes. ECLA are not available as yet on the Internet, nor are they incorporated yet into any databases that also have other substantial subject indexing, like DWPI.

U.S. Patent Classification. U.S. patent classes are the big issue because they are included in nearly every database that covers U.S. patents, either online or on the Internet. But in what form? Like ECLA, U.S. classes undergo steady-state revisions, and old patents are re-classified with new classes. Six times a year, the USPTO releases information on class changes and lists of patents reclassified.

As of now, the online U.S. patent databases—the CLAIMS and USP files, the full-text U.S. patent files on Dialog and STN—load current patent classifications at least once a year, and the STN full-text file loads them six times a year.

Of the Internet U.S. patent databases, the USPTO site loads current classifications; as do the two not-quite-Internet services, Corporate Intelligence and Chemical Patents Plus. All the others load only the classifications first assigned to patents, and the IBM site does not even let you search them directly. So you have the interesting situation that in databases in which free text and patent classifications are the only subject search parameters available, the patent classes are essentially useless unless you are searching in classes that have not changed since the 1970s.

ENHANCEMENTS

Internet databases in general mount patent data in the form in which they receive them. In contrast, some online database producers provide enhancements to their files over and above the specialized indexing and patent classifications. A few examples follow.

Text Enhancements. DWPI do not simply mount the titles and front-page abstracts that appear on patents, rather, they write their own high-quality technical abstracts and then write detailed patent titles from these. The IFI files start with the original U.S. patent titles but expand them.

Company Name Enhancements. Several of the online databases, the IFI files and DWPI, provide company codes which group variations on company names. IFI also standardizes company names. This, plus the ability to expand on company name indexes (discussed later), helps you avoid the guesswork involved when you search companies' patents on Internet sites. Of the Internet sites, only Patent Explorer includes company codes, the same ones that are in DWPI.

Patent Number/Filing Number Enhancements. IFI scans for and manually corrects errors in filing information in the data received from the USPTO. DWPI looks for nonconvention equivalents and adds them to the proper patent families.

Table 1. Current Capabilities of Internet Patent Resources^a

Capability	A	B	C	D	E	F	G	H
Current U.S. patent classifications?	Yes	No	Yes	No	No	No	No	Yes
Standardized company names/codes?	No	No	No	Yes	No	No	No	No
Sets of search results created/reusable?	No	No	No	Yes	No	No	Yes	No
Complex Boolean, multi-field searching available?	No	No	Yes	No	Yes	No	Yes	Yes
Indexes? (Of company and inventor names, patent classes, etc.)	No	No	No	No	No	No	No	No
Ability to generate patent statistics?	No	No	No	No	No	No	No	No
Search logic stored for future use?	No	No	Yes	Yes	No	Yes	No	No
Automated SDIs?	No	No	Yes	Yes	No	No	No	No
Forward/backward citation searches?	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Sorting of search results?	No	No	No	No	No	No	No	Some

^a A = Chemical Abstracts' Chemical Patents Plus; B = Community of Science site; C = Corporate Intelligence site; D = Derwent's Patent Explorer; E = IBM site; F = MicroPatent's PatentWeb; G = Questel-Orbit's QPAT-US; H = USPTO site.

SEARCH SYSTEM CAPABILITIES

Significant differences exist between Internet and online search systems. The apples-and-oranges analogy is especially relevant here, because most online systems are intended for professional searchers, whereas most Internet search systems are designed for end-users. They are usually menu-driven and limited in their capabilities and flexibility. The worst of the Internet search engines are frustratingly primitive. Some of them let you search only one term at a time. Some do not let you limit to fields. The best of the Internet search engines just begin to approach the basic capabilities of the online search engines. They provide fielded searching. They permit nested Boolean logic, proximity operators, single and multiple wild-card characters, and some numerical range searching for dates.

Sets. Most Internet engines do not let you create sets and combine the set numbers in subsequent searches. (QPAT-US and Patent Explorer are exceptions.) This capability is important capability for complex searches in which you want to put together different groupings of the sets you have created, eliminating duplicates between each group and all previous ones. One of my recent searches involved a four-step chemical reaction in which I looked at the intermediate chemicals as both products and reactants and then put together all possible reactant-product permutations. This would have challenged most Internet search engines.

Field Searching with Nested Boolean. The online systems also give you full Boolean capabilities combining multiple fields. In one of my recent searches I looked for separations technologies (for which I was searching both terminology and International Patent Classes) either in connection with a specific chemical, or as assigned to a particular company. On the Internet sites that provide search menus with separate search term boxes for separate patent fields, this would require multiple separate searches; and the Internet databases which automatically assign AND logic between fields would not permit OR logic between, for instance, text terms and patent classes. You can do this sort of search in one step only on the sites that provide an advanced search mode with nested Boolean and let you combine different fields in one search statement. So far, the IBM, QPAT-US, USPTO, and Corporate Intelligence sites let you do this.

Indexes. Online, you can expand on indexes. That is, you can see and choose from lists of search terms. You can choose whole ranges of company and inventor names, dates,

index terms, or patent classifications without having to key them in directly. Not only does this speed up entering of search terms, but it eliminates the need for guesswork: You can see and choose from all the forms in which that inventor's or company's name appears. None of the Internet patent sites has loaded indexes.

Linking. Some online databases have linking capabilities that give you greater precision in your searches of subject indexing. Link systems can be quite sophisticated. For instance, DWPI provides three levels of linking for the new polymer indexing that it introduced in 1993. The API indexing software generates permuted link sets of appropriately linked terms. Because the Internet databases do not have linkable indexing, the closest they come to this is proximity searching.

Patent Statistics. Most online systems have data ranking capabilities (RANK on Dialog, GET on Orbit, MEMSORT on Questel, SmartSelect on STN) that let you generate all sorts of patent statistics. For instance, you can do a general subject search and then produce a ranked list of the companies holding patents in the area. My colleagues and I have written several articles dealing with the sorts of statistical searches that you can do, and giving extensive caveats about what you might get.¹ Internet patent databases do not have these capabilities.

Saved Logic. Most online systems will let you save search logic, either temporarily or permanently. You can save "hedges," complex logic for subject concepts that you search frequently, and recall them any time by the names you have given them. Of the Internet databases, only the Patent Explorer, MicroPatent, and Corporate Intelligence sites have this capability.

SDIs. Most online systems also provide automated SDI capabilities. That is, they let you use saved search logic for ongoing searches run automatically against database updates to give new references as they are added to the database. Some online hosts will permit variable timing of SDIs; others are moving in that direction. So far, of the Internet databases, only Corporate Intelligence and Patent Explorer provide automatic SDIs. And most of the other sites do not permit the sort of fine time ranging that will let you do manual SDIs, that is, do a search and produce only new patent references. The MicroPatent site does permit week-by-week date ranging for the US, EPO, and PCT gazettes. It also offers a "This week's full text" free option for U.S. patents.

MULTIFILE/CROSSFILE SEARCHING

Probably the most important feature of online searching versus searching on the Internet is that the online hosts let you search multiple databases at once, and they let you cross patent search results between databases.

This has serious implications for patent searchers. As we all know, and as Stu Kaback and others have been pointing out since the early 1980s, no one database can provide a complete patent search, no matter how well indexed it is. Each database has its strong and weak points, and each will give unique good answers in a subject search. Patent searchers know how important it is that they use, not just one database, but all relevant databases for as nearly complete a search as possible.

This procedure has become much easier in recent years with the cross-file and multi-file search capabilities that most online hosts have developed. The best of them will let you search multiple databases at once, using each one's unique indexing, and download or print your search results in patent families neatly grouped across databases, with duplicate records eliminated or not as you choose. In some cases you can cross search results between databases at intermediate stages of your search, in effect combining the indexing of several databases—a process I have written about and dubbed “virtual file merging.”²

These capabilities do not exist on the Internet. In another article in this issue, Elvin Hoel discusses a do-it-yourself, labor-intensive cross-file search process with QPAT-US. In general, though, each Internet database is separate, and each must be searched separately. The closest thing to multiple databases is on the MicroPatent site, where you can search US, EPO, and PCT gazettes together. And even here, the system permits no identification or grouping of patent family members.

LINKS/CITATION SEARCHING

Probably the most useful feature on the Internet is the ability to jump from spot to spot with hypertext links. In all Internet patent databases you can go (relatively) quickly from the search screen to a help message and back again, from reference to reference, from a list of patents to individual full records and back again, and so on. In some, you can click on features within a reference to go other useful places. For instance, on the USPTO site you can click on a patent classification to go directly to its definition. On the IBM site, on the other hand, clicking on a patent classification will take you to a list of all patents in the database initially (not necessarily currently) classified there—the only form of class searching available on the IBM site for now. Also on the IBM site, a full-view display of search results includes some legal status codes, and clicking on these will take you to a list of legal status actions that the database covers.

All the Internet databases except the MicroPatent site use hypertext links to provide forward and backward citation search capabilities. That is, on the front page of any patent reference, you can click on individual patent numbers cited in that patent to go to their front pages if they are also in the database; or you can generate a list of more recent citing patents. QPAT-US takes this capability one step further, in a “network” citation search. This will give you a list of all

patents cited in the one you are examining, plus all more recent patents that cite any of them.

Online citation searching is not so visual, but is more effective. Suppose you know of several patents relevant to a subject that you are searching, for which you want to see more recent citing patents as a supplementary subject search. A quick online search lets you generate international patent families for all of them and search all their family members as cited patents, to find more recent patents from a number of countries (not just U.S.) that have cited any one of the family members of any one of your patents of interest. This provides a far more complete forward citation search than is possible on the Internet, and it gives you one final set of citing patents with no duplicate references.

TEXT CRUNCHING

Some companies now offer Internet-based services that do varieties of patent text-crunching. EDS's Shadow Patent Office, Manning & Napier's MAPIT, and QPAT-US look at text you input or text of whole patents or parts of patents, and generate lists of other patents in their database ranked by “similarity.” MAPIT, the most sophisticated of the three, will match up a set of your patents with a set of competitor patents in the same subject, pairing them off and ranking the pairs by similarity.³

How valuable these searches are is debatable. Sample searches that I have tried on SPO and QPAT-US have varied enormously in quality of search results, depending on the text I have asked them to examine. In any case, as a patent searcher I am reluctant to depend on computer algorithms whose actual processes are kept secret. But they could certainly be useful as supplementary subject searches. This sort of text crunching is not available in the online patent databases, for now.

WHAT TO DO WITH YOUR SEARCH RESULTS

Internet display systems have Internet-provided advantages. You can usually produce on your computer screen the full record that a particular database provides, up to the full patent text or image. If it is text, you can use Netscape or Explorer FIND capabilities to look for strings of characters, to help you browse the document. Some Internet databases highlight the terms that you have searched. All this is aimed at end users, who normally look through search results on the spot and make decisions about individual records, for instance whether to obtain full copies. You usually have no choice of the order in which your Internet search results are displayed. They are sometimes reverse-chronological but more often in some kind of relevance-ranked order. Edlyn Simmons discusses the value (or lack thereof) of patent relevance ranking in more detail in another article in this issue.

Most online systems do not provide the sorts of graphical displays of which the Internet is capable. However, they do provide a variety of standard display options or let searchers choose which parts of the records are of interest and then download or print exactly those parts for further examination. They also let the searcher choose, to some extent, in what order the records are displayed. Results are normally reverse-chronological, but sorting is possible on a variety of fields, for instance by company or inventor name. In short, display

options are designed less for the end user and more for search intermediaries who will tailor and edit search results to hand on to the search client.

However, the Internet does have online beat when it comes to looking at patents. In fact, a typical approach among patent information professionals is to search the indexed online databases and then look at their search results on Internet resources. They can also download patents from these Internet sources and incorporate them easily into search reports. On the IBM site you can look at actual full-patent images, including the drawings and chemical structures; and now you can gateway there from QPAT-US—and do so anonymously, which should address some concerns about confidentiality on the IBM site. So far, looking at full patent images is slow, especially compared with the speed of flipping through hard-copy patents in the shoes at the Patent Office. But it will speed up, and eventually we will have access to a virtual Patent Office at our fingertips.

The Internet also provides matchless convenience in ordering patents—as long as one of the sites has what you want. If you need a Romanian patent from 1969, or even a British patent from 1980, you'll have to cope with the British Library or some other international resource. But U.S., EPO, and PCT patents are all available (the U.S. clear back to 1790, from Corporate Intelligence) at the push of a button, so to speak.

CONCLUSION

What I have given you here is a snapshot in time, online and Internet resources as they existed on April 30, 1998. Many of the Internet and other patent services are gearing up to add enhancements and new capabilities, some of them this year; and new Internet sites, for instance of individual countries' patent offices, are appearing regularly.

I have not discussed cost issues so far, but certainly Internet patent databases are less expensive than the indexed online databases. At one extreme, you have a database that gives free searching of title, abstract, and claims text and free viewing of the full patent. At the other extreme, you have a database that costs \$100,000 or more per year just for the privilege of paying additional fairly steep online costs to search it and display patent references. But sometimes you get what you pay for.

Professional patent searchers are the only people who know both Internet and online databases in depth. We are

the only ones who understand their vital differences, who know when the more expensive indexed databases are essential to answer a patent question. End-users are very poor judges of the quality of their own searches; they cannot know what they did *not* get from an Internet search.

It is essential that we earn the trust of our own management. We must be allowed to judge what patent information resources we need access to. We must make management realize that our principle concern is the disastrous consequences to our companies of inadequate patent information.

URLs of the Internet Sites Discussed

Canadian Patent Office site:

<http://strategis.ic.gc.ca/cgi-bin/patent/searche.pl>

Chemical Patents Plus:

<http://casweb.cas.org/chempatplus>

Community of Science:

<http://patents.cos.com>

IBM site:

<http://www.patents.ibm.com/ibm.html>

MicroPatent site:

<http://www.micropat.com>

Patent Abstracts of Japan:

http://atlas.patent-inf.tu-ilmenau.de/brs/paj_eng.html

Patent Explorer:

<http://www.patentexplorer.com>

PCT Gazette site:

<http://pctgazette.wipo.int/>

QPAT-US:

<http://www.qpat.com>

USPTO site:

<http://www.uspto.gov>

REFERENCES AND NOTES

- (1) Lambert, N. Patent statistics searching: Practices and pitfalls. *Proceedings of the Montreux 1989 International Chemical Information Conference*; H. Collier, Ed., Springer Verlag: Heidelberg, 1989; pp 141–160; Simmons, E. M.; Lambert, N. Comparing grapes and watermelons. *ChemTech* **1993**, 23 (6), 51–59.
- (2) Lambert, N. Virtual file merging: A technique to enhance patent searches. *Database* **Oct 1997**, 20 (5), 36–44.
- (3) My next "Better Mousetrap" article (in *Searcher* magazine) will look at these text-crunching engines in more detail.

CI9801217