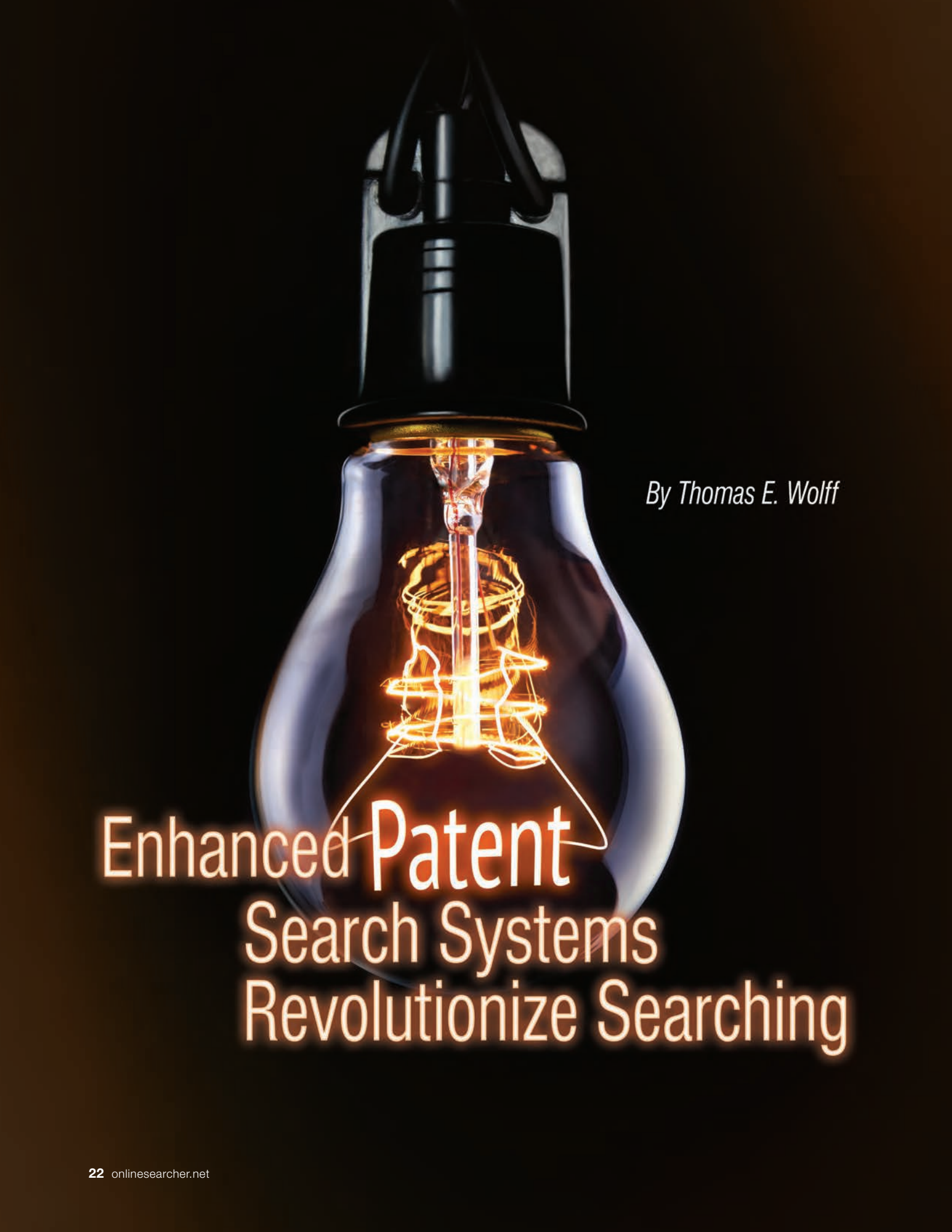


By Thomas E. Wolff

Enhanced Patent Search Systems Revolutionize Searching

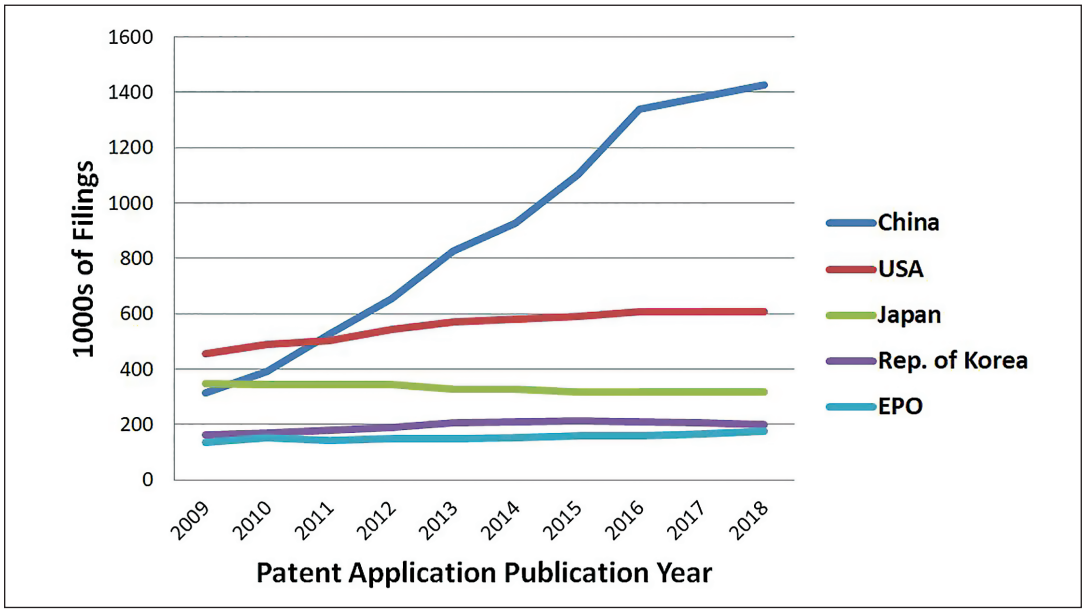
Patent database producers and their users continue to be challenged by the ever-increasing number of patent documents in a multitude of languages with various types of non-textual content. Increasingly sophisticated tools allow patent searchers to navigate the world's collection of patent documents and to be more efficient throughout the search process—from receipt of a search request to delivery of a final search report.

Professional patent searchers, in particular, have made gains from improvements in enhanced patent search systems from commercial sources. These improvements go well beyond what free patent databases offer and provide viable alternatives to expensive, value-added patent databases that index patent content.

A regular concern for database producers and professional searchers, expressed at technology conferences and in patent blogs, is the vastly increasing numbers of patent applications filed worldwide in many languages. Patent filings for the top five patent offices within the past decade, including filings made under the Patent Cooperation Treaty (PCT), show an enormous increase in filings from China, according to the World Intellectual Property Organization (WIPO) Statistics Data Center (wipo.int/ipstats). Plus, these filings could be made in any of three Asian languages (Chinese, Japanese, or Korean) or the three official European Patent Office (EPO) languages (English, French, and German). There are four additional official publication languages for PCT filings: Arabic, Portuguese, Russian, and Spanish.

In fact, the EPO and the WIPO International Bureau as receiving offices (wipo.int/pct/en/filing/filing.html) accept patent applications in almost any language as long as they are translated into one of the 10 official languages. Other patent offices take applications in their home languages without requiring translation into other languages.

>



Patent filings for the top five patent offices across the past decade, including filings made under the Patent Cooperation Treaty (PCT), showing the huge increase in filings from China.

PATENT DATABASES AND SEARCH SYSTEMS

The most economical way to deal with large patent datasets is to use automated processes for gathering and organizing patent information from sources worldwide. These processes scale up well as computing resources become faster and cheaper. The simplest patent search systems that facilitate access to the worldwide patent databases are produced by patent offices, search engine providers, and other organizations that choose to offer free access and limited search and output functionality. Examples include Google Patents (patents.google.com), Espacenet (worldwide.espacenet.com), and FreePatentsOnline.com. These free patent search systems are excellent resources for inventors, researchers, managers, and even patent agents and attorneys who do not need to carry out comprehensive searches to support legal or other high-stakes decisions.

Professional patent searchers, their employers, and clients rely on commercial patent search systems that offer many advanced features and facilitate efficient and effective patent searching. These enhanced patent search systems are supported by subscriptions and include Derwent Innovation (Clarivate Analytics), Orbit Intelligence (Questel), and PatBase (Minesoft). These enhanced patent search systems are the focus of this article.

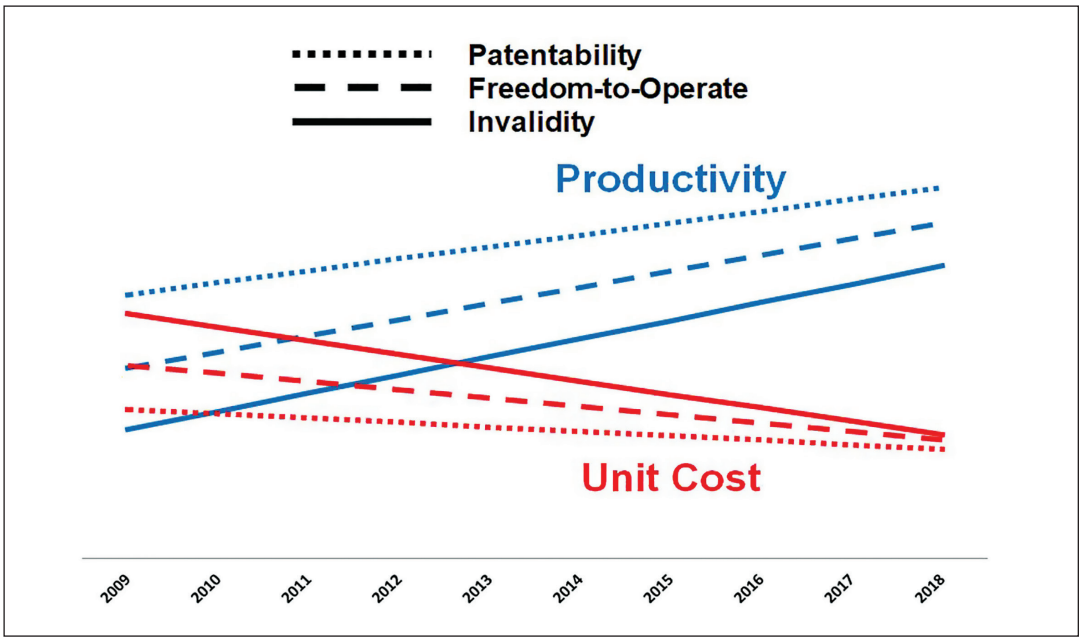
Many patent searchers, scientists, engineers, and attorneys continue to use older, value-added patent databases such as Chemical Abstracts' CAplus, Derwent World Patents Index, and EnCompass (formerly American Petroleum Institute) Patent Database. Other former favorites, such as the US CLAIMS database, are defunct. Producers of these value-added databases are particularly challenged by the increasing number of patents. They employ technical experts to index and abstract individual patents and applications and, in some cases, non-patent literature as well. However, these human efforts do not scale well, although they can be supplemented by computer-aided processes. Many users have dropped subscriptions to the value-added databases due to their relatively high subscription costs; the availability of non-subscriber, transactional access to much of the content; and the competitiveness of enhanced patent search systems.

How has this shift affected the ability of professional patent searchers to carry out effective patentability, invalidity, and freedom-to-operate (FTO) searches? What are the factors that affect the productivity of individual patent searchers like me? What conclusions can be drawn about how these trends have affected patent search quality?

SEARCH PRODUCTIVITY AND UNIT COST IMPROVEMENTS

My personal productivity, measured in the number of search reports that I can produce in a week, month, or year, has increased, even while the amount of patent data that I search and review has grown. Correspondingly, the average amount of time that I spend on each search project, from request intake and client interview to issuing the final report, has decreased significantly.

To confirm this, I analyzed the amount of time per search (unit cost per search) and productivity (efficiency corresponding to searches per unit time) for my work during the past decade. I graphed my unit cost data and productivity (smoothed by linear regression) for patentability, invalidity, and FTO search reports. The slopes of the search data are intended to be representative of the actual data, but the y-intercepts are absent in order to maintain business confidentiality.



A graph of my unit cost data and productivity (smoothed by linear regression) for patentability, invalidity, and FTO search reports

Advances in enhanced patent search systems have revolutionized how patent searchers work. Enhanced patent search systems allow for a start-to-finish search process.

These results correspond to 25–50% reductions in unit costs, leading to productivity improvements of 50–100% across the past decade. Improvements in unit cost and productivity were significantly greater for invalidity and FTO searching than for patentability searching.

SEARCHER EXPERIENCE AND SEARCH TECHNOLOGY

Searchers gain proficiency across time with search tools and resources and with their clients' technologies. However, repeat work from loyal clients on technologies that match my expertise cannot account for more than a modest portion of the reduction in unit costs of my searches of the past decade. The most significant changes in my personal growth took place more than 10 years ago when I started to document my processes on invalidity and FTO searching [1, 2, 3].

Most of the reduction in cost for my search services across time is attributable to technology improvements in the search tools I use. To remain competitive, vendors naturally make changes to improve their search products and develop new functionality that takes advantage of new computing technology.

These improvements affect various types of searches in proportion to how labor-intensive they are. In general, FTO and invalidity searches are more labor-intensive than patentability searches because they involve detailed consideration of large numbers of patent documents. FTO searching involves review of claims of granted patents and pending patent applications in the countries or regions of interest to the client. The cost of such searches is mitigated somewhat because FTO searching is restricted to patent documents applied for within the past 20 or so years, corresponding to potential enforceability.

Invalidity searching is more intense than patentability searching because the searcher must scour patent specifications (and literature articles, especially when readily available and reasonably priced in full text) to find disclosure of matters claimed by the patent that is the target of the invalidity search. The cost of invalidity searching may be lessened for older target patents because the search covers patents and non-patent literature that were filed or published before the filing date of the target patent document. In these cases, searchers can help attorneys evaluate results of invalidity and FTO searches by pointing out specific claims, specification text, or both in the search reports. In contrast, patentability searches are less rigorous and require less labor to review prior art and report findings. They are usually early-round novelty searches of proposed inventions. The potential penalty to the client of unreported prior art is less compared to invalidity and FTO matters.

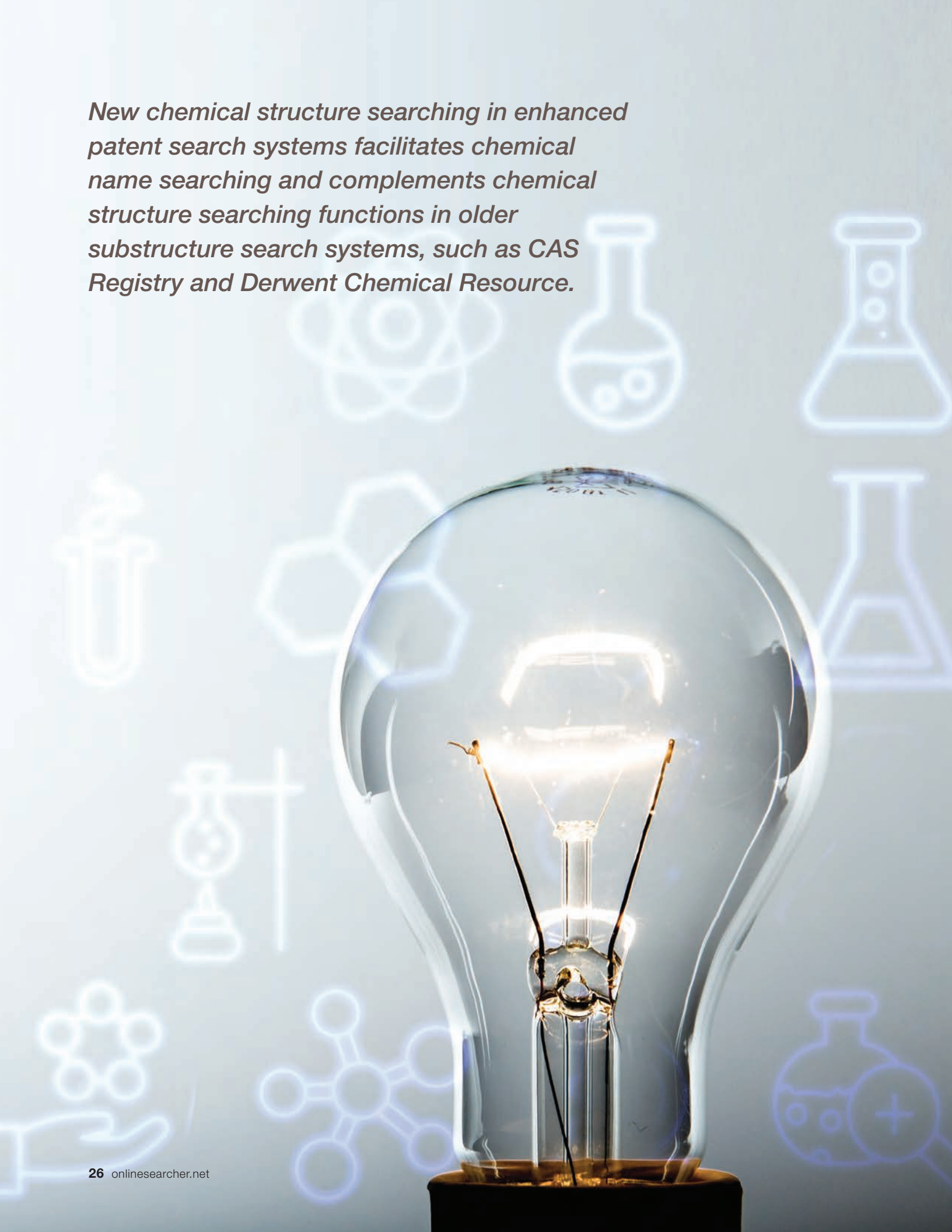
VALUE-ADDED DATABASES

Advances in enhanced patent search systems have revolutionized how patent searchers work. Enhanced patent search systems allow for a start-to-finish search process. Searchers can cycle through search, retrieve, review, learn, revise strategy, and repeat easily and repeatedly. They can review candidate references in detail during the course of the search, which has the biggest impact on FTO and invalidity searching.

Searchers no longer need to rely on value-added databases, except in some specific technology areas, to find manageable

>

New chemical structure searching in enhanced patent search systems facilitates chemical name searching and complements chemical structure searching functions in older substructure search systems, such as CAS Registry and Derwent Chemical Resource.



sets of patent documents or patent families for further personal review and evaluation by researchers and patent attorneys or agents. There is no need to order patent documents separately and interrupt the search process. In effect, the searcher becomes the real-time expert and diminishes the need to outsource to other technical experts to read, interpret, abstract, and index original hardcopy patent documents. This is the basis for expensive subscriptions or bulk-usage agreements.

Many companies, such as my former employer and many of my current corporate clients, have significantly reduced or eliminated their budgets for value-added databases. Unfortunately, they have also reduced the numbers of company-employed searchers because their technical and legal staff, management, and business leaders believe they can match the use of value-added databases with the patent systems they have available to them. They are missing several key issues when using only free patent databases: Value-added databases are still important for supplementing enhanced patent search systems in some technologies and are invaluable in others. Experienced professional searchers are still critical because they make the best use of value-added databases and enhanced patent search systems in order to deliver the search reports that clients need.

Value-added databases still excel over enhanced patent search systems for searching information that is not disclosed well for easy retrieval by current patent search systems. These include chemical image or line structures and substructures, Markush queries, polymers, protein and nucleotide sequences, and other graphic figures. Database indexers convert this information into content that can be searched in the value-added database systems.

The absence of such facilities in enhanced patent search systems is partially mitigated by the integration of patent classification and citation searching as alternative search approaches. Note that value-added non-patent literature databases, such as Chemical Abstracts CAplus and the corresponding SciFinder product, continue to be of good value because there are few public-domain, free-access, or value-added literature databases comparable to those in the patent space.

The coverage and indexing policies of value-added databases are not sufficient for all types of searches. The value-added database producers have policies that instruct indexers on which and how many patent documents from a patent family to index and on what content to index. These databases generally index records based on only one or a limited number of family members. Indexers may focus particularly on claims, on novel concepts, or on examples rather than index the full specification.

This indexing may not be thorough enough for FTO searches that relate to specific patent authorities (countries or regions) or target subject matter buried in claims. The indexing may not be deep enough to find information critical to success in an invalidity search. In such cases, searches employ a value-added database to identify patent families that should be evaluated further by transferring patent numbers to patent search systems or other sources of patent documents.

I find that value-added database indexing is most useful for patentability searching because the indexing frequently points to novel content. Most searchers would consider valued-added database searches as complementary to, but not substitutes for, full-text searching, especially for FTO and invalidity searching. This is particularly true, for example, for Chemical Abstracts CAplus and Derwent World Patents Index that remain critical for exemplary FTO searches on chemicals and polymers. There are similar enhanced databases in the biological and pharmaceutical areas.

TECHNOLOGY'S IMPACT ON THE FOUR PHASES OF SEARCHING

Many improvements in enhanced patent search systems relate particularly to these specific phases of the search process: 1) search strategy development, 2) strategy implementation, 3) records review, and 4) report preparation.

1. Search strategy development

I have become more productive in this phase by improving interviewing skills and developing search targets and goals as described in "Freedom-to-Operate Search Strategy: Hitting the Target" [3]. I often carry out limited initial searching in enhanced patent search systems prior to a client interview to build my knowledge of the technology and to characterize the scope and cost of a search. The process usually involves exploring patent classifications, which is facilitated by tools such as the patent Classification Explorer in PatBase. Alternatively, I may also use simple search tools such as Google Patent search or semantic search functions in enhanced patent search systems, particularly to avoid missing obvious references that clients may find on their own. Searchers never want to miss obvious references.

2. Strategy implementation

Enhanced patent search systems such as Derwent Innovation, Orbit Intelligence, and PatBase allow searchers to search and review original documents in ways that were previously impossible. New chemical structure searching in enhanced patent search systems facilitates chemical name searching and complements chemical structure searching functions in older substructure search systems, such as CAS Registry and Derwent Chemical Resource. Searchers input chemical structures, then the enhanced patent search system converts the query into text search terms in the background.

I usually supplement these structure search functions with my own text term searching, particularly for invalidity and FTO searches. This chemical search function works for documents in all of the principal languages and can be directed to claims or to full patent specifications. I would like to be able to direct structure searching to the combination of title, abstract, and claims, as I do in most text-term searching. As computing hardware and software improve, I look forward to be being able to search by chemical structure for highly posted, simple chemicals that currently exceed system limits and to use similar structure-based queries for polymers, Markush claims, and

>

biological sequences. Developers should also add reactant, product, or other descriptive roles for these compounds and materials, which are available in some value-added databases.

Some enhanced patent search systems facilitate searching in non-English patent documents by accommodating non-Latin language search terms and providing for search in high-quality machine translations of the original documents. I anticipate systems will fully incorporate machine translations into the main search and review functions. They would replace use of separate search strings targeting machine translations or separate review screens for translated documents when search algorithms and database hardware improve adequately.

Patent citation searching and semantic or natural language searching are usually integrated in enhanced patent search systems. I conduct citation searching regularly because it is independent of terminology and patent classification. I have not found semantic or natural language searching to be particularly useful, but it is easy to use. Early search systems such as PatCafe that implemented latent semantic analysis a decade ago did not provide an appropriate means of reviewing results that we take for granted today in enhanced patent databases systems.

3. Records review

The combined benefits of advanced, color-coded, hit-term highlighting and review of machine translations of non-English documents are a big factor enabling nearly complete and efficient review of candidate patent documents. Hit-term highlighting, which may apply to full specifications and to keyword in context (KWIC) displays, enables easy screening and detailed review of documents.

Some systems have progressed beyond automated or manually designated lists of key terms for highlighting to allow for Boolean logic to designate the relationships between the terms to be highlighted. For example, in a search on coatings or treatments involving silica, the search logic might include `(silica or silicon dioxide or silicic oxide or silicon iv oxide) near (coat or treat)`. With this improved functionality, the search terms are highlighted only when they are in proximity rather than in every occurrence in the KWIC or full displays.

Improved access to patent legal status is another helpful aid for FTO searching. Some systems have integrated simple visual indicators for being "alive" (potentially enforceable, perhaps in green) or "dead" (no longer enforceable, perhaps in red) for individual patent family members. At this time, most systems do not allow searching these legal status attributes. When it becomes available, searchers will have to be cautious because of the potential for "dead" documents being revived, such as by the payment of late maintenance fees, litigation, or the possibility of overcoming other delays in prosecution. In theory, one could craft a search to try to eliminate from consideration older documents that may truly be stone dead. In addition, I frequently report "dead" patent documents in FTO reports to allow client attorneys to look for safe harbor references and deal with uncertainty in legal status.

4. Report preparation

Clients appreciate getting selected claims and excerpts of specifications that support the selection of a patent document for further consideration. This used to be done by multiple copy-and-paste steps during the review process. Now some search systems have added convenient, time-saving functions that allow searchers to create notes associated with patent records or patent documents. These notes may then be exported with each selected patent record and processed for inclusion in the final report. I provide links in spreadsheets to full patent records and to legal status summaries when provided by patent search systems. I would like to be able to provide links directly to particular sections of patent text, such as claims or examples, and to have clients see records with highlighted terms.

THE FUTURE FOR PATENT SEARCHING

My ability to carry out higher-quality searches and provide better search reports, as well as my overall increased productivity, correlates well with increased reliance on enhanced patent search systems relative to value-added databases. Enhanced patent search system developers deserve tremendous credit for providing professional patent searchers with tools that make our jobs economically competitive and enjoyable, particularly in the face of ever-increasing numbers and languages of patent documents. They have done this even as they have designed their search systems to accommodate casual searchers as well. In particular, they have developed very critical interfaces that have significantly decreased the amount of time needed to analyze individual patent documents—claims, specifications, or both—for invalidity and FTO searching. For example, the ability to focus on search terms with advanced hit-term highlighting and the availability of more and better machine translations mean that searchers are less likely to miss important content and can make better decisions about accepting or rejecting patent documents.

Further improvements will undoubtedly come with advances in computing resources. I look forward to complete integration of machine translations in search and review processes. Machine translations would be treated exactly as original language documents and would be used by searchers as an equal information source. Searchers would still have to recognize that machine translations are not perfect and do not have the legal status of certified expert translations. Patent searchers should always be conservative in their judgments in order to provide all appropriate information to their clients.

Enhanced patent search systems will continue to create new alternatives to standard text term searching. Chemical structure searching should be followed by similar functionality, enabling searching for Markush claims, for addition, condensation and natural polymers, and for biological sequences. Information must be efficiently retrievable from tables; the relationship between elements in table rows is valuable and should be searchable with appropriate proximity operators applicable within individual or nearby rows and columns. These systems should provide access to textual and non-textual information in figures

Enhanced patent search systems will continue to create new alternatives to standard text term searching.

and images. Searchers would like to search for specific visual information contained in utility patents just as they might in design patents. While inventors are expected to describe figures within the specification, there is always other information that could be learned from the images themselves.

These systems should add numerical property searching by indexing numbers and their units within specifications, thereby providing full numerical searching. Such searching is currently available in some value-added databases and could be implemented in enhanced patent search systems much in the way chemical structure searching was. For example, this functionality would enable searching for a boiling point range in either Fahrenheit or Celsius, or particle size range in microns or nanometers. In each case, the property should be searchable in proximity to a target substance or material with the required numeric property.

PREDICTING REVOLUTIONARY CHANGES

Up to this point, I have been describing improvements that evolve in the course of product development. Revolutionary changes are more difficult to predict. I recall first hearing about natural language processing a couple of decades ago and automated chemical structure assignment in patent documents a few years ago. The latter quickly became implemented by commercial patent search systems, whereas the former has yet to make its mark.

The next extension of natural language processing seems to be the application of artificial intelligence (AI). IBM calls this augmented intelligence, because such enhancements aim to work along with searchers during the course of their work. This was a big topic at the PIUG 2018 Annual Conference and elsewhere. My concern is that AI seems to offer yet another way to find candidate references. That could be useful, particularly to casual searchers and some patent examiners, but is not likely to make a significant difference in how professional patent searchers do their work.

I understand that there is interest in using AI processes to improve the patent review process. Systems are being developed that would learn from the searcher's acceptance and rejection of candidate records, presumably by updating the ranking of candidate record sets. I leave it to others to program how an AI system could determine the rationale that a searcher used for rejecting a patent, particularly in FTO or invalidity cases which require detailed evaluations of claims or specifications. In addition, I would be concerned about using AI systems that were not integrated with patent search systems from strategy development to report generation. Developers of new systems will make inroads with professional patent searchers only if they recognize that we rely on such well-integrated systems.

References

1. "Freedom-to-Operate Patent Searching: My Six Basic Rules," *Searcher: The Magazine for Database Professionals,* 16(5) (May 2008): pp. 34–39; wolffinfo.com/docs/FTO_Patent_Searching_Wolff_in_Searcher_Magazine.pdf

2. "What's So Special About Patent Invalidity Searching?" *ONLINE*, 36(4) (July/August 2012): pp. 28–33; wolffinfo.com/docs/Invalidity_Searching_Wolff_in_ONLINE_Magazine_Jul-Aug-2012.pdf

3. "Freedom-to-Operate Search Strategy: Hitting the Target," *Online Searcher*, 42(4) (July/August 2018): pp. 31–42; wolffinfo.com/docs/FTO_Search_Strategy_Wolff_in_Online_Searcher_Magazine.pdf

Acknowledgments

The author offers his sincere thanks to Sandra Unger, Rahman Hyatt, Mark Frentrup, and Barbara Staman Wolff for their insights and feedback on this paper.

Thomas E. Wolff *(tom@wolffinfo.com; www.wolffinfo.com) formed Wolff Information Consulting, LLC in 2006 to provide technical and patent information services, specializing in the chemical arts, on a contract basis, following a career at Amoco Chemical Co., BP, Innovene, and Ineos. He has 2 U.S. patents and is a registered patent agent with the USPTO. Wolff is certified by the International Standards Board (ISBQPIP; qpip.org/about) as a Qualified Patent Information Professional (QPIP).*

Comments? Email the editor-in-chief (marydee@xmission.com).