



Lora Burgess

# Curation reveals hidden information for more effective IP searching

Lora Burgess, Chemical Abstract Services (CAS), discusses the ways curated content could improve your patent search by making it easier, more precise and less time consuming.

**B**usiness decisions are only as good as the information that guides them. Nowhere is that more true than in the realm of intellectual property (IP): success or failure can hinge on an invention's novelty, the strength of a patent when challenged, or infringement on an existing claim. IP professionals must identify how and where they can efficiently find the high quality information needed to make critical decisions. There is a wealth of options from which to source information to support intellectual property decisions, and they all fit roughly into one of two categories: full-text or curated content. Patent attorneys will always need to look at the full-text of key documents for final analysis, but curated content can provide a significant efficiency boost to help pinpoint those key documents quickly while ensuring that important sources are not overlooked.

## Full-text options

The websites of major global patent authorities like USPTO, EPO and WIPO, as well as some information providers like Google Patents, provide access to the full-text of granted patents and published patent applications for selected countries. Finding documents from most full-text sources is based on keyword searching - the user enters one or more keywords, and the search engine retrieves patents containing those keywords. However,

terminology in patents can vary and is often obscure. To be successful, a searcher must choose the right keywords for a query or risk missing relevant patents. In addition, the patent text contains many words leading to over-retrieval of false hits. As a result, obtaining precise (yet complete) search results through full-text patent searching presents a challenge.

Once a relevant patent is found, getting to the pertinent information within the patent presents an additional barrier. Full-text patents provide the complete, unabridged documents and all of the information they contain. But with patents reaching into hundreds and sometimes thousands of pages, it is burdensome and time-consuming to sort through all this information.

The language in a patent adds further complexity; it is often deliberately vague to maximize the breadth of the patent protection. Given the above, searchers must consider the practicality of combing through full-text literature page by page as a preliminary assessment of relevance.

## Curated content databases

Thankfully, there is an alternative. There are organizations dedicated to getting information seekers to the answers they need more efficiently through curated content. In the context of database content, curation entails the intellectual analysis of primary sources by highly trained experts in relevant fields such as the sciences, technology, and engineering. Typically specializing in a given subject or discipline, these experts then index documents for enhanced searchability and aggregate their analyses into databases made available to searchers through software and web applications. When considering curated content, searchers must balance comprehensive search retrieval and opportunities for increased precision to arrive at the best results before performing post-search review and analysis.

## Résumé

### Lora Burgess

Lora has been working at CAS for over 16 years as a trainer and intellectual property specialist. After working in the corporate world for a number of years, she returned to school to get a BS in chemistry from Texas State University, and a Master's in chemistry from Johns Hopkins University. She worked as a search expert and trainer at the USPTO before joining CAS.



### Balancing comprehensive search retrieval and precision

Intellectual analysis of curated databases allows for comprehensive search results, so searchers can be confident they're considering all relevant information. Within a patent, something as simple as a kitchen table might be described as "a rectangular planar surface with four supports"; never even using the word "table" in the description. By indexing this information, curated content providers help searchers identify complex information in patents by translating concepts into clear and searchable index terms.

Patent information is sometimes presented in the literature using graphical representations, such as drawings, chemical structures, or biological sequences. The indexing process takes the concepts in drawings and tables and renders them into text. Conversely, substance information is often portrayed in lengthy text descriptions and R-group tables. Some curated databases turn these easy-to-miss descriptors into searchable chemical structure diagrams. Both of these curation approaches make the content easier to search, find and digest.

Intellectually-enhanced titles, abstracts, supplementary terms, and standardization of new terminology increase comprehensiveness in searches. For example, an original patent titled "Method for Synthesis of Aromatic Amine" is given the following intellectually enhanced title in a curated database:

*Process for synthesis of substituted secondary amines via condensation of aniline with aryl halides with a palladium catalyst and (t-Bu)<sub>3</sub>P as a ligand as an electroluminescence source for display devices.*

Company name variations and mergers are included in searches as well. For example, a search on Genentech will show you different company names used by Genentech to file patents over the years: Genentech Inc., USA; F Hoffmann La Roche Ag; Genentech Inc.; F Hoffmann La Roche AG, Switz., etc.

Although a comprehensive search of raw data alone will not produce the level of accuracy needed, indexing will fortunately enhance search precision. For example, a functional role can be assigned to chemical substance index terms. "Diagnostic use" would be assigned to a substance used as a tool for diagnostic purposes (to identify a disease, pathogen, or injury) in laboratory tests or imaging techniques. Expert linking of chemical substances with their function in a particular document

enables a much higher degree of search precision to retrieve only the most relevant answers. In this way, indexing helps searchers strike the appropriate balance between precision and comprehensiveness.

Another aspect of precise information retrieval in a curated content database is the use of controlled vocabulary and thesauri. These lists of standardized terminology, words, or phrases used for indexing and information retrieval are enhanced by natural language selected from the patent document. New terminologies, frequently occurring terms, concepts and synonyms are regularly added by content experts so the thesauri stay up-to-date. For example, a search on lymphocytes would include synonyms such as immunocyte, lymphoid cell, leukocyte, B cell, plasma cell, natural killer cell, T cell, and so forth to capture more relevant answers.

STN®, a patent search platform popular among IP professionals, provides access to numerous curated patent and bibliographic databases and their related thesauri. These thesauri can be used to find terms to narrow or broaden your searches, depending on your need. For instance, a search on nanodevices retrieves over 4,000 references; the CAplus<sup>SM</sup> thesaurus can be used to narrow your focus to pharmaceutical nanocapsules to extract the most relevant results (Figure 1 below).

Remarking on the services CAS provides to customers, Christine McCue, vice-president of marketing at CAS, explained: "CAS has been at the forefront of curating scientific information for decades. Our goal is to make information retrieval more efficient, and the literature more accessible for searchers, expediting their scientific and IP research and, by extension, accelerating scientific innovation." Thanks to services like those provided by CAS, and curated intellectual analysis by subject matter experts, searchers can save considerable time narrowing in on documents of key interest to their search queries before they begin detailed analysis.

The screenshot shows the CAplus Controlled Terms (Lexicon) interface. The search term is 'nanodevices'. The results are displayed in a hierarchical tree structure with expandable/collapsible icons and counts for each term.

Term	Count
<b>Nanodevices</b>	<b>4070</b>
Broader	3
Narrower	31
Synonyms	4
Related	3
Electric nanogenerators	30
Molecular electronic devices	2680
Molecular photoelectric devices	32
Nanoactuators	94
Nanocapsules	1385
Pharmaceutical nanocapsules	2383
Nanoelectrodes	1035
Nanofilters	708
Nanofiltration membranes	2727
Nanomachines	2227
Molecular shuttles	195
Molecular switches	2930
Nanoelectromechanical systems	1506
Nanomotors	187
Molecular motors	604
Nanoneedles	104
Nanoreactors	834
Nanoscale semiconductor devices	1837
Nanosensors	2502

The screenshot displays the PatentPak interface. At the top, there are navigation controls for 'PAGE 301', 'ZOOM', and 'DOWNLOAD PDF'. The main content area shows a patent document with a chemical structure and text. A 'Highlighted Location Marker' points to a specific part of the text. A 'Marked Patent' box is also visible. On the left, a 'Key Substances Sidebar' lists three substances with their CAS numbers and chemical structures: CAS RN 1786400-23-4, CAS RN 1786400-24-5, and CAS RN 58803-45-2. Each entry includes an 'Analyze It' button and a 'Page' indicator.

## Post-search review and analysis

Earlier, we discussed some challenges presented by searching full-text documents via patent office websites and Web search engines. Using curated databases can help you more quickly identify important documents.

Once you have found a document of interest, some curated databases even provide links directly to the full-text of the literature so you can obtain it immediately once you've pinpointed a result of interest. Related references and citations are usually linked with the search results as well for convenience.

The standardization of patent assignee information and terminology also streamlines analysis of trends, key players in a technology area and emerging technologies.

Until recently, few tools existed to help searchers locate relevant details within a patent, so there was often no alternative to scanning each page for important information. PatentPak™, a recently-introduced solution from CAS, addresses this problem for the chemical information space.

Leveraging indexing performed by CAS scientists, PatentPak provides direct links to the location of key substances in the full-text document.

One click takes the searcher from the CAS databases to the exact page in the patent where the substance is described, be it in a table, chemical structure, paragraph or other format.

Thanks to new time-saving solutions like PatentPak, analysis of full-text patents is becoming a less onerous task.

## A real-life example

Perhaps the best way to appreciate the value curated content can add to IP research is through a real world example. Take a patentability search for sensors that measure gases in blood or blood plasma. A searcher might start in freely available full-text sources, but that can quickly become overwhelming.

A search of “gas sensor blood plasma” (before priority of January 01, 2011) on Google Patents finds over 28,000 documents, many not relevant to the topic of interest. Even after sorting by relevance, answers like “Fluoropolymeric substrates with metallized surfaces and methods for producing the same” (US5945486A), “Inductive Plasma Processor Including Current Sensor for Plasma Excitation Coil” (WO2002080220 A1) and “Improved Apparatus and Method for Aliquotting Blood Serum and Blood Plasma” (WO1992019949 A1) are returned near the top of the results list. These answers mention gas passageways and gas supplies, but are not focused on gas sensors.

In contrast, retrieval in the human-curated CAPLUS database returns approximately 300 precise references about sensors used to measure gases in blood. These include “Blood oxygen saturation detection device and patient monitoring wireless device comprising same” (KR2016004163 A), “Diagnosis of Plasmodium infection based on detection of volatile organic compounds” (WO2015077843 A1) and “Apparatus and method for monitoring of an extracorporeal blood circuit for detection of air bubbles” (DE102012024341 B3). This enhanced precision is made possible expressly because of the intellectual indexing provided by content area experts.

## Conclusion

IP professionals need to make confident decisions on patentability, freedom-to-operate and validity searches. The indexing efforts of technology experts that create curated content can be leveraged to improve the comprehensiveness and efficiency of your patent searching. Indexing and intellectual analysis results in greater precision and more reliable results. Coupled with the availability of tools to analyze the information more rapidly, these benefits make curated content an essential resource for many experienced IP searchers' workflows. The next time you're confronted with half a million patents in a search result, consider the time-savings afforded by curated content.