11-2007

Patent Coverage in Scientific and Technical Research Databases

Jan Comfort
Clemson University, comforj@clemson.edu

Leena Lalwani
University of Michigan, llalwani@umich.edu

Patent Coverage in Scientific and Technical Research Databases

Jan Comfort
Government Information Librarian
R. M. Cooper Library, Clemson University
comforj@clemson.edu

L**eena Lalwani**
Coordinator for Arts and Engineering Collections
Art, Architecture and Engineering Library, University of Michigan
llalwani@umich.edu

Abstract

Patents play an important role in the research process. Yet, most recent articles about patent searching and sources of patent information focus on tools and techniques used by information professionals, not researchers. This article will compare the features of a number of research databases that provide patent information. The analysis will include a chart listing cost, dates of coverage, search strategies, and the benefits and limitations of each.

Keywords: Science and Technology Databases, Patent Searching, State-of-the-Art Patent Searches, Search Strategies, Information Needs of Scientists, Engineering Village, Google Scholar, SciFinder Scholar, Scitopia.org, ISI Web of Knowledge

1. Introduction

Patents are an important component of research databases. A survey of the patent literature shows that most articles focus on patent-specific databases. However, Bonnie Snow wrote two articles in 1989 comparing the coverage of patents in non-patent databases in the fields of bioscience, and food, agriculture and environment. She highlights why patent literature is daunting to researchers, and why subject-oriented databases can be an important tool in acquiring knowledge about patents in their field. (Snow, 1989a; Snow, 1989b)

In a 1998 article on patent coverage in scientific databases, the authors point out that researchers are more likely to be familiar with literature published in their field than in patent information published by the PTO. They conclude that because of limited coverage of patents, searchers should consider searching in subject-oriented databases as the first step in doing a more comprehensive search in a patent-specific database. (Carpenter & Hart, 1998)

Author Nancy Lambert offers “A Snapshot in Time” in her 1999 article. At the time this article was written, online sources of patent information (such as DIALOG, STN, and Questel-Orbit) were the primary providers of patent information. Internet databases were relatively new on the scene: The full-text database at uspto.gov had just been introduced and it covered only 1976 to date. Other internet databases were also less comprehensive than the online sources mentioned above, but they were affordable, and offered many convenient features. (Lambert, 1999)

In a recent article, authors Mechtild and Wolfgang Stock compare the main providers of online intellectual property information (Thomson DIALOG, Questel-Orbit, and STN International) and ESP@CENET, a web-based provider. They list several ways that these databases can be improved, in particular by adding more complete information. They conclude that more research is needed in describing and analyzing other information providers, as well as understanding the needs of researchers and information professionals. (Stock & Stock, 2006)

2. Purpose

A review of the most recent patent searching literature documents changes in how patent information is presented in specialized patent databases. We were interested to see if there were any analogous changes in patent coverage in subject-specific research databases. We decided to take a closer look at some of the most widely available research databases, to compare their features and benefits. The purpose of our analysis is to evaluate the depth of coverage of patents in several scientific databases used by researchers for state-of-the-art searches. For our purposes, a state-of-the-art patent search is defined as a comprehensive search of all information about a technical field, including patent and non-patent literature. (Hunt, Nguyen, & Rodgers, 2007) A list comparing these features is included in Appendix A. To be included in our analysis, the database had to meet several criteria:

- First, it must cover a wide range of scientific information in all formats, not just patents. Therefore, specialized patent databases (such as the databases available at the USPTO website, MicroPatent, Lexis/Nexis Patents, and Google Patent Search) were not considered. There is no question about the value of these resources, but because of their specialized nature, they play a different role in a state-of-the-art search. Nor have we included ESP@CENET, the patent database from the European Patent Office. Although it contains one million non-patent literature references, the articles are not easily searchable.

- Next, we focused on databases that are available directly to end-users. We looked at research databases in science and technology that are either subscribed by institutions or available at no cost. We have not included services from vendors like DIALOG or STN, which are generally used or mediated by an information professional, and are charged by the use rather than by subscription.

- Finally, the database must be noted in *Search Templates* on the United States Patent and Trademark Office (USPTO) website, located at http://www.uspto.gov/web/patents/searchtemplates/searchtemplates.htm. Search Templates are formatted lists of sources, chosen based upon input from patent examiners and other searchers at the USPTO, and reflect their institutional knowledge of the most relevant prior art search sources. Scitopia is a new database and hence not in Search templates.

- A number of databases did not meet all of the above criteria, but because of their prominence in a particular field of study, we have provided information about their coverage of patents. These databases are listed in Appendix B.

3. Databases

3.1 Engineering Village

Engineering Village from Elsevier is a web-based information service providing access to Compendex® (1969-date); Engineering Index Backfile (1888-1968); Geobase; INSPEC; NTIS; Referex; and granted patents and patent applications from the United States Patent and Trademark Office (USPTO) and the European Patent Office (EPO). Each of these databases is priced individually based on other purchases on Engineering Village, size of institution and type of institution. Patent coverage includes: US Patents (from 1790-date); published US Patent Applications (2001-date); and documents from the EPO and comprises 10 million documents.

Figure 1. © Elsevier. Used with permission.

Benefits

Using Quick Search, you can search all the databases together or limit by specific database. Once the search results are displayed, you can refine them by using the faceted column on the right-hand menu bar, such as Database (Compendex, US Patents or EP Patents), Author/Inventor, Author Affiliation/Assignee, Controlled Vocabulary, Classification Code, Country, Document Type or Year. When you search only patent databases, you can narrow your results to patent-specific fields like Database (US or EP), Patent Type (application or granted patent), Inventor, Assignee, USPC (United States Classification Code), ECLA (European Classification) Code, IPC (International Patent Classification) Code, Country and Year. You can search within results by entering a term in a search box at the bottom of the facets section to limit results without doing an additional search. To keep track of your searches, you can easily view search history and combine results. You can also sort results by relevance, date or author. You can save results to a disc, export to a bibliographic management tool, print or e-mail. You can save searches, or create folders to keep your searches organized. You can also sign up for free weekly e-mail alerts, or an RSS feed to automatically receive new records (up to 400) when the database is

updated. Full-text links take you directly to a PDF of the image of subscribed articles. Direct links are also provided to cited articles, cited patents and citing articles.

Using Expert Search, you can perform even more sophisticated searches by limiting searches to a number of additional fields. Those fields related to patents are: Patent Application Date, Patent Examiner, Patent Issue Date, Patent Number, Patent Citation Index, Patent Application Country, Patent Authority Code, Patent Filing Date, Patent Application Number, and Patent Attorney Name.

Limitations
The greatest disadvantage of this database is the inability to search by current USPC and current ECLA. You can search using assigned classification code, but patents prior to 1933 do not have assigned classification printed on the patent and hence will not be included. Even in the faceted results the option for narrowing is assigned classification and not current US or ELCA classification. This is problematic, as experienced searchers use current USPC or current ECLA because it incorporates the changes to the classification number over time. Bibliographic information for older US patents is added via scanned OCR (Optical Character Recognition) which has many errors. A look at US patent 222,134 for *Improvement in Roach Traps* illustrates the problems associated with OCR.

3.2 Google Scholar
Google Scholar is a service provided by Google™ allowing key word searching across many disciplines. It includes peer-reviewed papers, theses, books, abstracts; and articles from academic publishers, professional societies, preprint repositories, and universities. This database is available for free. Google Scholar accesses patents that are in Google Patent Search, Patent Storm, WIPO Patentscope, some national patent offices, FreePatentsOnline, etc. An exhaustive list of patent sources covered in Google Scholar is not available at this time. Information is extracted from above-mentioned databases and information on coverage of some follows: Google Patent Search covers U.S. patents from 1790 through mid-2006. At this time it does not have a schedule for regular updates. Coverage from FreePatentsOnline includes US patents issued since 1976, published US applications since 2001, and European patents since 2000. Patent Storm covers patents issued from 1990 to present, and is updated weekly. Each of these databases can be searched individually.

Benefits
Results are sorted by relevance, based on patent title and inventor. In some cases different patent numbers are grouped together, aiding in identifying related patents. European patents from FreePatentsOnline link directly to esp@cenet, which is a bonus. Both Patent Storm and FreePatentsOnline have added subject indexing by categories. The Google patent viewer is free, doesn't involve a download, and permits downloading an entire patent in PDF. Records contain hot links to Abstracts, Drawings, Description, and Claims, allowing users to easily go to relevant parts of the document; there is also a link to view the entire patent at uspto.gov. Hot links for Current U.S. Classification link directly to the Classification Definitions on uspto.gov, enabling users to easily reach the proper place to do a comprehensive classification search. Google Patent Search only includes US patents, so using Google Scholar could provide a means for identifying relevant foreign patents. When viewing articles, users who are affiliated with a university (and

have set their preferences in Google Scholar) can link to available full-text resources without the need of performing additional searches in the Library's catalog. When patents are cited in an article, they are highlighted in the references, providing easy access to citations.

Limitations

Unfortunately, there is no way to remove duplicates. Many of the records from Google Patent Search were duplicated from other providers. Results from FreePatentsOnline include lots of advertising, and will only link to full-text patents in PDF after registering for a free account. Adding the search term "–freepatentsonline.com" will exclude those original records, although they will still be included in the groupings from Google Patent Search. This search strategy will remove some duplicates, but it is not intuitive for most users. There is no way to search within results to quickly identify all hits on a particular patent number or author/inventor without doing additional searching. Although having results sorted by relevance is generally acceptable, there is no way to sort or limit by document type. It is not possible to limit by patent number, inventor, assignee, classification or date.

3.3 ISI Web of Knowledge™

ISI Web of Knowledge™, produced by Thomson Scientific, brings together journal articles, patents, websites, conference proceedings and open access material in all fields of science. ISI Web of Knowledge covers Web of Science® and other ISI citation databases; BIOSIS Previews®; Biological Abstracts®; CAB Abstracts®; FSTA (Food Science & Technology Abstracts®); Inspec®; MEDLINE®; and analytical and bibliographic management tools. Each of these components is priced individually based on other purchases on ISI Web of Knowledge. In addition, ISI Web of Knowledge provides access to Derwent Innovations Index℠ (DII), which includes Derwent's World Patents Index® with the Derwent Patent Citation Index™. This index covers over 16 million patents from 40 worldwide patent-issuing authorities. Patent coverage goes back to 1963, and cited references are from 1973.

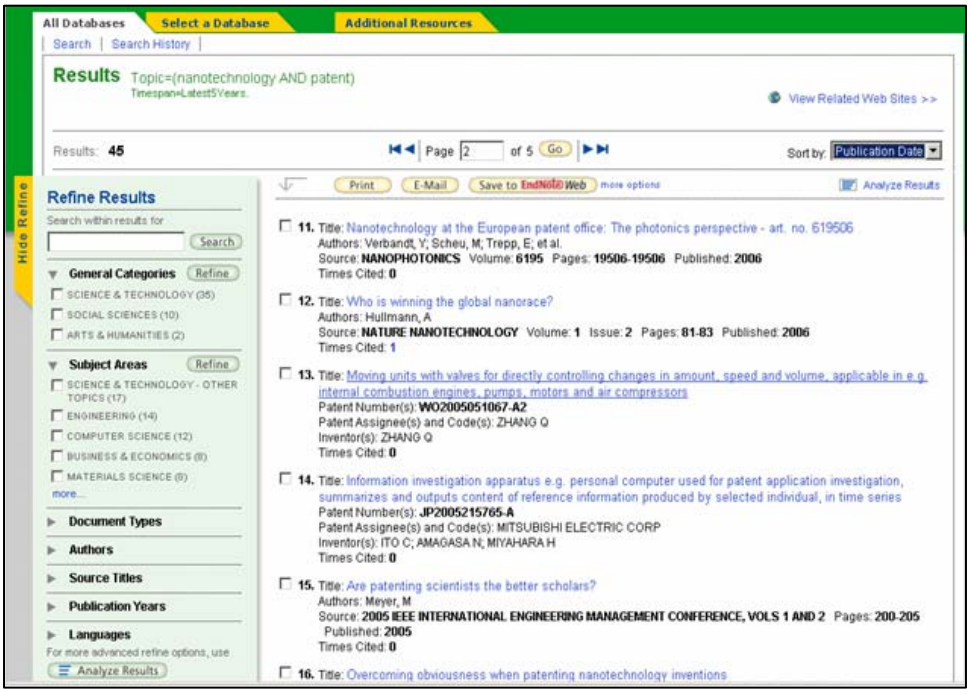

Figure 2. © Thomson Scientific. Used with permission.

Benefits

Use the CrossSearch feature to search across all databases. Search choices are limited to TOPIC and AUTHOR/INVENTOR, but once you have search results you can filter by database. Choose DII to retrieve patent results. DII adds a considerable value to this database, including a vendor-supplied document summary with a more descriptive title, understandable abstract, chemical structure and patent family data (showing relationships between patents that have been filed with one application in multiple countries). In addition, Derwent assigns a unique Derwent Class and Derwent Code, facilitating the identification of related patents. All of these fields are searchable, as are the Assignee Names, Assignee Codes, Inventors, Patent Number, and International Patent Classification Codes. Results can also be analyzed by all searchable fields. DII also has a specialized Compound Search, allowing advanced searching for chemicals.

Click on Cited Patent Search to search by Cited Patent Number, Cited Assignee, Cited Inventor, and Cited Derwent Primary Accession Number. Results can be sorted by: Latest Date, Inventor, Publication Date, Assignee Name, Assignee Code, Times Cited, and Derwent Class Code. Clicking on the patent title brings up a brief abstract, with tables of patent family information and links to other documents (IPC, Derwent Manual Code, etc.)

Click on Advanced Search to search by field. All of the fields mentioned above are searchable. Additional features include search aids which provide the following browsable lists: Inventor Index, Assignee and Code List, International Patent Classification List, Derwent Class Code List, Derwent Manual Code List, Cited Inventor Index, and Cited Assignee and Code List. Codes bring similar items together. e.g. The assignee code pulls together patents issued to all subsidiaries of a company along with the parent company.

Common features of all the databases in ISI Web of Knowledge [SM] include the ability to view search history, combine sets, and create alerts or RSS feeds. Full-text links take you directly to PDF of the image of subscribed articles. ISI Web of Knowledge [SM] unveiled a new interface at the end of August 2007. The CrossSearch feature (allowing simultaneous searching of several databases) had major enhancements, but the DII native interface remained the same.

Limitations

Patent coverage in this database begins with 1963, so it is not a good resource for historical and genealogical research. Since this database adds so much value, there is a time lag of about 20 days before information available. The coverage is selective, even in its strongest subject areas. Citing patents are not necessarily the same as the citing patents listed in the USPTO PatFT database. Derwent Codes are helpful for use within this database, but are meaningless when searching other databases.

3.4 SciFinder Scholar™

SciFinder Scholar™ is produced by Chemical Abstracts Service (CAS®), a division of the American Chemical Society. It is the leading service providing information on chemical literature, and includes journal articles, book chapters, patents, conference proceedings, technical reports, dissertations, and substance and structure information. In addition to the CAplus, CASREACT, and Registry databases from CAS, SciFinder Scholar™ searches Medline®-a medical literature database produced by the National Library of Medicine. Patent coverage is part of the database and cost to the database varies depending on the size and type of institution. CAS has recently improved their patent coverage: adding records for chemical patents from the United States back to 1790, patents covering biological sequences, and expanding access to foreign patents. It covers patents from 50 patenting authorities worldwide. Patents now make up 16% of records in Chemical Abstracts.

Benefits

Click on “Explore,” to search by Author Name, CA Section Title, Company/Organization, Database, Document Type, Index Term, Journal Name, Language, Publication Year, and Supplementary Term. Click on *Locate* to search for a CAS Registry number, or a particular article by keyword, author, or journal name; or search by patent number, inventor last name or assignee. In addition, you can also search by chemical structures or biological sequences, substructures and reactions. Results can be refined (to include or exclude records) by: Research Topic, Company Name, Author Name, Publication Year, Document Type, Language, or Database. When articles are marked, click on “Get Related” to view: Cited References, Citing References, Substances, Reactions, or e-Science (information from the web). There are direct links to Full-text articles from subscribed journals. Bibliographic information about a patent, including patent family data can also be easily accessed. Each US patent record is linked to the uspto.gov website, requiring a TIFF viewer. EP documents link to esp@cenet. You can save citations in different formats like ASCII, RTF, Tagged, tab delimited, Answer Keys, and quoted format to load into bibliographic management software. However, this feature is not as easy to use as in Engineering Village or Web of Knowledge. Results can also be e-mailed. The 2007 version has a categorizing function available, which is similar to the faceting function in Engineering Village.

Limitations
SciFinder Scholar™ only includes a subset of chemistry patents based on IPC code and USPC. A list of classifications covered can be found at http://www.cas.org/expertise/cascontent/caplus/patcoverage/index.html. Cited and citing references links are buried under "get related." There is no way to create an alert or an RSS feed. In order to use the database, you need to install the SciFinder Scholar client software on your computer: it is not a web-based platform. Since there are many value-added features to this database, there is a time lag of 4-6 weeks from the time a patent issues to the time the value-added information regarding the patent is added to the database.

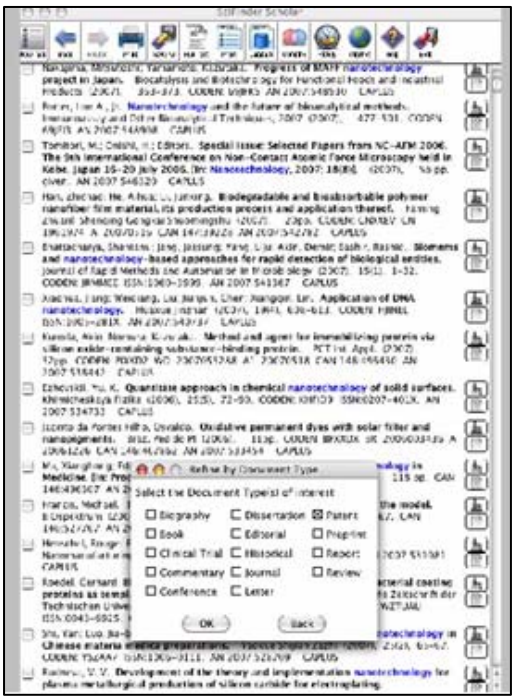

Figure 3. © CAS. Used with permission.

3.5 Scitopia
Scitopia is a new free internet portal for scientific information created by 14 science and technology societies. It includes more than 3 million peer-reviewed journal articles, technical papers, and patents spanning 150 years. IEEE covers applications and issued patents from USPTO, EPO and JPO. It is a very new platform. It appears that USPTO patents are from1976 and forward, but coverage is not clear from the documentation. It is available at scitopia.org.

Benefits
All standard features are available, including Boolean searching, phrase searching, and truncation. Advanced search permits fielded searching: Full Record, Title, Author, Abstract, Affiliation, and Date Range. You can refine searches further by limiting publications to one or more of the societies. Results can be saved to *My Articles*, which is available until you close the browser. Items in *My Articles* can be sent via e-mail or RSS feed. The patent coverage is very current, as it gets its data via live feed from the respective patent offices.

Limitations

Be patient. It takes a little time for Scitopia to search each of the sources. Only the first 100 results from each publisher are included. Abstracts of articles are linked to full-text, but viewing requires a subscription, or purchase of individual articles. Searching does not work properly all of the time, especially author searching. It is not possible to limit searches to only patents: results are categorized, and patents are identified in a tab. Classification searching is not possible at this time, nor does the database include information on cited or citing patents. US Patents link to the USPTO website, requiring a TIFF viewer. EPO and JPO patents link to esp@cenet, and therefore provide PDF images.

3.6 Scopus®

Scopus® is a bibliographic database that covers citations and abstracts from journals, conference proceedings, trade publications, book series, and references added to these citations. In addition, Scopus also covers web sources, which include 18 million patents from the World Intellectual Property Organization (WIPO), EPO, USPTO, and JPO. Access to patents is included in the cost of the database and the cost varies on the type and size of the institution. Scirus – available at www.scirus.com – is a free interface that has the same patent coverage as the Scopus database.

Benefits

To search for patents, perform an author or a keyword search and view the results in the patents tab on the results screen. You can refine results by patent office by adding a keyword in the Refine results box. You can save results to a disc or to a bibliographic management tool, print, or e-mail. Alerts and RSS feeds are also available.

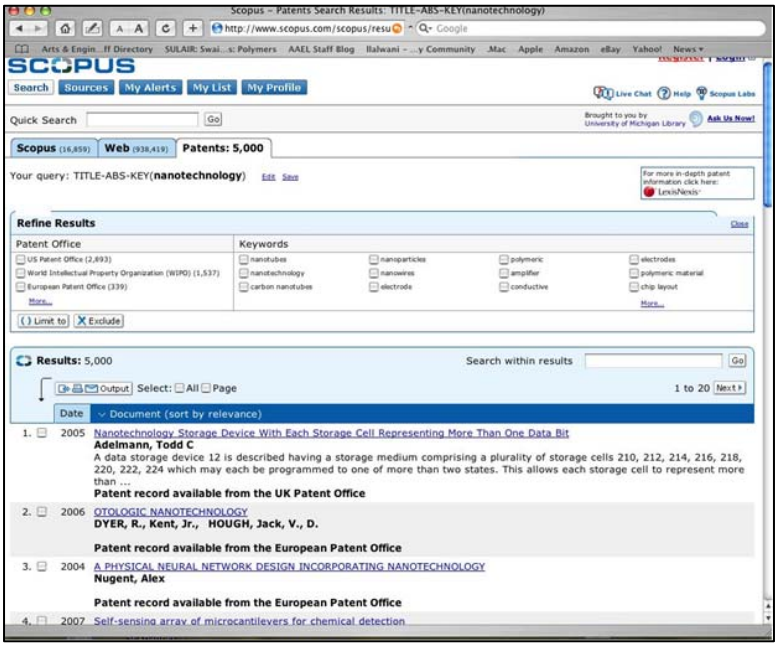

Figure 4. © Elsevier Scopus. Used with permission.

Limitations
None of the patent-specific fields (classification number, assignee, etc.) can be searched, so results cannot be restricted to patents. Older bibliographic information for US patents is added via scanned OCR, which contains many errors. Another look at US patent 222,134 illustrates the problems associated with OCR. Search results of US patents are linked to the USPTO site, requiring a TIFF viewer. There are no links to cited patents or citing patents. Scirus, the free database from Elsevier is not a good alternative. It is slow and clunky, and returns many errors.

4. Conclusion

Each of the databases we reviewed has a place in a state-of-the-art search, and each has unique features and benefits. Searches in these databases can identify relevant patents and journal articles, and can help researchers choose keywords that can be used for searching in patent-specific databases. But no single database can be depended upon to give complete patent results. When performing a state-of-the-art patent search, researchers should consult one or more of the comprehensive subject databases: Engineering Village, SciFinder Scholar, or Web of Knowledge. But they shouldn't stop there. Because of limitations in subject scope, search capabilities, and dates of coverage, they won't know what they've missed. Researchers should build on these results by searching in a patent-specific database, to ensure complete results.

5. References

Carpenter, B., & Hart, J. L. (1998). Jump starting the patent search process by using subject-oriented databases. *Database Magazine, 21*(6), 20.

Hunt, D., Nguyen, L., & Rodgers, M. (2007). *Patent searching : Tools & techniques*. Hoboken, NJ: Wiley.

Lambert, N. (1999). Patents on the internet versus patents online: A snapshot in time. *Journal of chemical information and computer sciences, 39*(3), 448-452.

Snow, B. (1989a). Patents in non-patent databases - bioscience specialty files. *Database, 12*(5), 41-&.

Snow, B. (1989b). Patents in non-patent databases - food, agriculture and environment files. *Database, 12*(6), 115-119.

Stock, M., & Stock, W. G. (2006). Intellectual property information: A comparative analysis of main information providers. *Journal of the American Society for Information Science and Technology, 57*(13), 1794-1803.

Appendix A. Databases Reviewed

	Engineering Village	Google Scholar	ISI Web of Knowledge	SciFinder Scholar	Scitopia.org	Scopus
Dates of Coverage						
US patents	1790+	1790+	1963+	1907+	1976+	1960+
US patent applications	2001+	2001+	2001+	2001+	2001+	2001+
EP patents and applications	esp@cenet	esp@cenet	1963+	1907+	esp@cenet	esp@cenet
Other	No	Yes	Yes	Yes	No	Yes
Subject coverage	Complete	Complete	Selective	Chemistry	Complete	Complete
Publication lag	No	No	Yes	Yes	No	No
Thesaurus	Yes	No	Yes	Yes	No	No
Cost	Yes	Free	Yes	Yes	Free	Yes
Special Features						
Search speed	Good	Good	Good	Good	Fair	Good
Truncation	Yes	No	Yes	Yes	Yes	Yes
Auto stemming	Yes	Yes	Yes	Yes	No	Yes
Phrase searching	Yes	Yes	Yes	Yes	Yes	Yes
Limit to patents	Yes	No	Yes	Yes	No	No
Browse indexes	Yes	No	Yes	Yes (Some)	No	No
Boolean	Yes	Limited	Yes	Yes	Yes	Yes
Proximity	Yes	No	Yes	Yes	Yes	Yes
Patent family	Indirectly	Indirectly	Yes	Yes	Indirectly	Indirectly
Fielded Search	Yes	Limited	Yes	Yes	Limited	Limited
Patent number	Yes	No	Yes	Yes	No	No
Author/inventor	Yes	No	Yes	Yes	No	Yes
Company (Assignee)	Yes	No	Yes	Yes	No	No
Citation Search						

Cited references	Yes	Yes	Yes	Yes	No	No
Citing references	Yes	No	Yes	Yes	No	No
Classification Search						
USPC current	No	Yes	Yes	No	No	No
USPC issued	Yes	No	Yes	No	No	No
IPC	Yes	No	Yes	No	No	No
ECLA	Yes	No	Yes	No	No	No
Post-search						
Sort results	Yes	No	Yes	Yes	Limited	No
Search within results	Yes	No	Yes	Yes	Limited	Yes
Limit results	Yes	Yes	Yes	Yes	Limited	Yes
Combine searches	Yes	No	Yes	Yes	No	Yes
View search history	Yes	No	Yes	Yes	No	No
Remove duplicates	Yes	No	Yes	Yes	No	No
Save results	Yes	No	Yes	Limited	No	Yes
E-mail results	Yes	No	Yes	Yes	Yes	Yes
Create alerts	Yes	No	Yes	Yes	No	Yes
RSS feed	Yes	No	Yes	Yes	Yes	Yes
Create reports	No	No	Yes	Yes	No	No
Link to full image	Yes	Yes (TIFF for US)	Yes	Yes (TIFF for US)	Yes (TIFF for US)	Yes
Export to citation manager	Yes	Yes	Yes	No	No	Yes

Appendix B. Other Databases

Database	Patent Search Fields	Notes
Biosis Previews 1986-present	DT=Patent PA=Patent Assignee PC=Patent Country PD=Patent Date PN=Patent Number	Searches the Official Gazette. Not available in Biosis Previews or Archive.
Agricola 1979-July, 2002 with selected earlier patents	Patent in so	Selective coverage of plant patents.
CAB Direct 1932-1988	In Advanced Search and Expert Search,patent in Publication Type field.	
CSA Materials Research Database with METADEX 1966-date	PA=Patent Application Data PC=Patent Country PN=Patent Number PR=Patent Priority Data PT=Publication Type (Patent)	
Food Science and Technology Abstracts 1990-present	DT=patent PA=Patent Assignee PAA=Patent Assignee Address PC=Patent Country PD=Patent Date PPD=Priority Patent Date PPR=Priority Patent	For archived records, the document type is journal-article-patent.
Inspec 1968-1976	DT=Patent NU=Patent Publication Numbers	
MicroPatent Materials Patents 1996-present	PA=Patent Application Data PN=Patent Number PR=Patent Priority Data PT=Publication Type (Patent) AP=Patent Applicant IC=International Class IN=Inventor UC=US Class	Monthly Updates of US, EP and PCT patents in fields relating to materials science.

Pirabase 1975-present	In Advanced Search, limit to patent as Document Type.	Covers all areas of packaging, pulp and paper, printing, publishing, and imaging.
Polymer Library 1978-1980, 1994-present	PA=Patent Application Data PN=Patent Number PR=Patent Priority Data PT=Publication Type (Patent)	Fomerly RAPRA abstracts.
PubMed 1980-present, selective coverage	patent[properties] and <keyword>	Nucleotide and Protein databases only.
Textile Technology Index 1944-present	Advanced search, limit to patent as Document Type.	