

Commentary

Patents—Overlooked Source of Information

Introduction

Patents are a rich, though often overlooked, source of information on technological advances. They can provide the knowledge of the current state of the art, enable one to watch the development of the industry as a whole or the important activities of competitors, assess the uniqueness of one's own invention, or find solutions to specific technological problems. A series of historical patents published in this journal provide insight into the early years of thermal spray technology.

When one looks in the references of technical papers, patents are quite rarely cited. A possible reason for their lower popularity may be the less straightforward “legalese” language and often a long time period between the application filing and publication of the granted patent. However, with proper tools that empower “information mining,” the usefulness of data contained in patents can be greatly increased. This article briefly introduces the reader to the capabilities of the on-line patent database provided by Delphion Inc. (www.delphion.com) and provides examples of ways to get the most information from this valuable resource.



Jiri Matejcek



Jan Ilavsky

Patent Searching

Delphion^[1] offers numerous search options and capabilities, as well as broad coverage. Currently available collections are: United States, European, German, and Japanese (abstracts only), WIPO (World Intellectual Property Organization), and INPADOC (International Patent Documentation File). For the U.S., European, and German collections, one can choose either or both of granted patents and published applications. The coverage dates differ from collection to collection, but can go back as far as the 1960s.

One can perform a quick search by keyword or phrase, choose among the above collections, and select whether the search terms should be on the front pages or full text of the documents. For quick access to a specific patent, its number and country can be entered directly.

Boolean search (Fig. 1) lets the user match keywords to specific fields in the patent (e.g., the title, inventor, assignee, abstract, claims, etc.) and join these fields by Boolean operators (and, or, and not); the keywords are entered in a predefined form. Additionally, the range of dates of coverage can be specified. When performing the search, one can also select the fields to be displayed in the grouped results, the number of results per page and how they will be sorted (e.g., by relevancy score, title, assignee, publication date, etc.).

Advanced query lets the user perform a search with all the above options by entering the search parameters (keywords, dates) and operators as a text in a specific query language. The advantage is that the parameters can be combined and nested without limitations and even complex logical structure can be built as desired. By enabling the searches to be highly specific, the user can spend less time looking for information and more time acting on it.

Besides the Boolean operators mentioned above, one can use the location operator <in> to specify that the keyword should appear in a given field. For example, the following query:

```
(spray AND apparatus) <in> TI AND Plasmatechnik <in> PA
```

will yield patents that contain the words “spray” and “apparatus” in the patent title and that are assigned to Plasmatechnik (the abbreviations TI and PA stand for Title and Patent Applicant/Assignee, respectively, but full words can be used as well). Proximity operator <near/n> selects documents containing two or more search terms within *n* words of each other.

When search terms are entered together without a Boolean operator, they are considered a phrase. By default, word stemming is applied so that documents containing keywords having the same word stem as the search term are retrieved—for example, entering `spray` will result in `spray`, `spraying`, `sprayed`, `sprayable`. In contrast, entering “`spray`” will yield just `spray`. Wildcards are also possible, with `?` representing exactly one character, and `*` representing zero or an unlimited number of characters. The wildcards can be placed at the beginning, in the middle, or at the end of a word.

The application or publication dates can be limited either from below or above or both (when looking for patents older than or newer than a specified date or from a specific time interval) using the operators `<`, `>`, `=` and their combination. The following example:

```
(spray torch <in> TI) AND (PD>1988-12-31 AND PD<1990-01-01)
```

will retrieve documents containing the phrase “`spray torch`” in their title and published in the entire year 1989.

Fig. 1 Example of Delphion Boolean search. Several typical search fields are preselected, and examples of entries are automatically provided on the right.

When many documents are expected, different weights can be assigned to individual keywords, and several sorting schemes can be chosen. Also, alternative language can be specified for patents where it is available.

All the options and procedures of the advanced search are explained in detail on the Delphion website; here only a brief overview illustrating its capabilities is given.

Besides patents, non-patent prior art can be searched through forms linked to other databases.

The search queries can be saved for later use—then rerun, edited, copied. Also, alerts can be set up, so that the user is notified when the collection is updated.

Data Mining

Finding the relevant patent(s) is only a start ... In most cases, a number of related patents are found, and they now need to be mined and data in usable form need to be extracted. The following illustrates how to take advantage of the database structure and how to get more information from other related documents and from the patent itself.

The search results can be sorted by any field that was chosen for display—e.g., by relevancy to the search terms, date, assignee, and so forth. They can be saved as “work files” for later use, edited (e.g., new dates can be entered for the same set of keywords), annotated, and shared with other people (see example below).

A summarizing snapshot can be generated using the same categories. This is a table showing a number of patents from the search results for each category. For example, a snapshot summarized by assignee will show companies/institutions that hold the most patents and thus can serve as indicator of the most active “players” in the field. Similarly, summarization by year can show the development history of that field, indicate whether it is growing and how fast, and so forth. A number of snapshots can be displayed at a time.

The patents can also be clustered based on the keywords. Linguistic and relational analysis is applied to the title and abstract. A visual

map of the clusters can be generated, as well as new work files from each cluster. This tool can help identify the most relevant patents and examine the relationships between patents.

The patent display page (here called the Integrated View) contains a wealth of information. Besides the standard data presented on the original document, a variety of information from different sources is linked from this page.

The Derwent Title contains a concise, descriptive, English-language title written by Derwent experts to highlight the content and novelty of the invention disclosed in the patent. This one sentence is sometimes more informative than the full original abstract, which can be overly verbose and generic. More explanation can be obtained on the document kind (whether it is a granted patent or application, and so forth—these categories vary by country). The text of all claims, summarizing description and drawing description can be expanded, and a PDF version (with a choice of resolution) of the entire patent can be obtained. Another link is provided to more information about the patenting company.

Related documents can be also accessed from the front page. More patents on a similar topic can be reached by clicking on the classification links: IPC Code, U.S. Class, and Field of Search (in a U.S. patent, again this varies by country). A link is also provided to other patents from the same institution/company. Also, patent family members, that is, patents filed with different patenting authorities that refer to the same invention, can be displayed. Direct links are provided to backward references (patents and applications cited as references by this patent or application), forward references (patents or applications that cite this patent as a reference), and other references (nonpatent prior art that this patent references, for example, journal articles and company reports).

The Citation Link tool will display the reference relationships (“citation tree”) graphically for easier analysis. This is useful for assessing the role of a given patent in the industry—what was the prior state, what are the following developments, who’s working in the field, and what impact did the patent have. A variety of display attributes can be controlled for easier analysis.

Exporting the Information

The data from the search results can be exported in a variety of formats for use by other applications, for example, spreadsheet editors, database software, and text editors. The formats include CSV, Tagged, RIS, Derwent, and XML. The fields to be included in the output can be selected.

Example

As an illustrating example, a set of patents from the fourth quarter of 2003, included in the patent overview for the Volume 13(2) June 2004 issue, is publicly accessible as a work file at the following address:

https://www2.delphion.com/public_list?wref=6758291&did=bXoLY01oZbM%3d%0a

Users not already registered with Delphion will be asked to register before viewing the work file.

Other Sources

The search capabilities of several other major resources are described here.

U.S. Patent and Trademark Office (USPTO)^[2] offers a separate search for patent applications (published since 2001) and granted patents (full-text since 1976, full-page images since 1790). The quick search option lets the user enter two different keywords or phrases, match them either to a specific field or have them searched in all fields, and connect these two pairs by a Boolean operator. Year coverage can be also selected from three options: 1976 to present, 1790 to present and 1790 to 1975; the 1790-1975 range is searchable only by patent number and classification. Similar to search using Delphion, advanced search permits one to form a complex query by entering the keywords and field codes (out of 31 available fields), Boolean operators, and parentheses in a text box. For the dates, one can specify either an exact date or a range of dates or use a wildcard instead of a day. This example:

`ttl/spray and isd/1/$/1998`

will return patents that contain the word “spray” in the title and that were issued in January 1998. The three year coverage options mentioned previously are available in addition to the text query. Quick patent number search is also available; unlike in the other databases, the number has to be entered with commas (e.g., 5,146,634).

The patent front page presents the basic patent information, as well as all claims, invention description, and drawing descriptions as a text, with the keywords highlighted. Links to backward and forward references (among U.S. patents) are provided.

The original documents can be viewed as TIFF images of single pages; this requires a special plug-in for the Internet browser.

Apart from patents, trademarks can be searched from a separate database.^[3]

Espacenet^[4] is a patent database maintained by the European Patent Office.^[5] It can be accessed through gateways in national languages of the member states. The patent collection in the database, though, is not limited to Europe. One has the choice of performing the search in one of the four collections: worldwide, EPO, WIPO, and Japanese abstracts. In the quick search, one can also select whether the keywords should be (a) in the patent title or abstract or (b) persons or organizations. In the advanced search, keywords can be entered in a form for specific fields (ten fields are available), which are by default joined by AND. In both quick and advanced search, more keywords can be entered in one form field and Boolean operators can be used as well. While the “form” version of

the advanced search is not as flexible as the “text” query at Delphion and USPTO, it is simpler and easier to use, as there is no special syntax to be learned. Two more quick options are available: search by number and by classification.

The amount of information displayed on the patent front page varies with the source. It always presents the patent bibliographic data. Where available, an abstract is also displayed (sometimes, an abstract of a correspondent document is displayed) as well as tab links to description and claims. If the invention was patented in other countries as well, a link named “also published as” is provided to these documents. Another link is provided to other patents in a family (patents related via priority). Cited documents are only available for European and PCT (WO) applications. The original documents can be viewed as PDF files of single pages.

The World Intellectual Property Organization (WIPO)^[6] maintains a database of international applications.^[7] These documents are somewhat different than national applications or granted patents. The Patent Cooperation Treaty (PCT) lets the inventor file a single international patent application with WIPO. WIPO performs a preexamination, and the applicant can then request patent protection in as many signatory states as needed.

Several search options are again available, starting with a simple search form with a choice of searching for either any or all of the keywords entered or exact phrase. Structured search allows one to enter the keywords into form fields designated to patent fields (more than 20 available) and connect them with five different operators (and, or, not, xor, near). Advanced search allows one to form a text query with a syntax similar to those described previously; date range selection, nesting, truncation proximity (within five words), and phrase searching are available. In addition, selection of front pages/full text and a specific week is available for these two types of searches. For users who create an account, a history is kept of the last 20 searches performed. These can be added to new searches. Several display and results ranking options are also available.

The original documents are linked to Espacenet.^[4] Besides the PCT documents, trademark, industrial design, and emblem collections are searchable.

Links to other national and international patent offices and similar institutions, patent databases, and other information sources can be found, for example, on the EPO links page.^[8]

References

1. Delphion, <http://www.delphion.com>.
2. USPTO Patents and Applications Database, <http://www.uspto.gov/patft/index.html>.
3. USPTO Trademark Electronic Search System, http://tess2.uspto.gov/bin/gate.exe?f=login&p_lang=english&p_d=trmk.
4. Espacenet, <http://www.espacenet.com/>.
5. European Patent Office, <http://www.european-patent-office.org/>.
6. World Intellectual Property Organization, <http://www.wipo.int>.
7. WIPO Intellectual Property Digital Library, <http://www.wipo.int/ipdl/en/>.
8. EPO—Patent information on the Internet, <http://www.european-patent-office.org/online/>.

Jiri Matejcek

Institute of Plasma Physics
Prague, 18221, Czech Republic

Jan Ilavsky

Argonne National Laboratory
Argonne, IL 60439