World Patent Information 31 (2009) 278–284

Contents lists available at ScienceDirect

World Patent Information

journal homepage: www.elsevier.com/locate/worpatin

Full-text prior art and chemical structure searching in e-journals and on the internet – A patent information professional's perspective ☆

Maik Annies *

Syngenta Crop Protection Münchwilen AG, Schaffhauserstrasse, CH-4332 Stein, Switzerland

ARTICLE INFO

Keywords:
Non-patent literature
Full-text searching
Prior art
Patentability
Novelty
Validity
E-journals
Chemical structure searching
Chemical identifier
InChI
Chemical data mining

ABSTRACT

In depth analysis of non-patent literature prior art is a crucial step in checking patentability of new inventions and validity of competitor's patents, since by patent law relevant subject matter disclosed in non-patent literature is as important as any patent document. E-journal articles, as well as any scientific and technical information published on the web are an important source of prior art that is very often insufficiently covered and indexed by commercial databases. This article reviews search and display capabilities of e-journal search sites of different publishers and hosts, as well as their value for full-text prior art analysis to enhance retrieval from commercial databases. Moreover, current developments and future prospects of chemical structure searching both in e-journals and on the internet are discussed.

1. Introduction

Commercial exploitation and economic success of new technologies and inventions are critically dependent on obtaining and maintaining patent protection. Additionally, it is important to avoid infringement of competitor's patents while new products are brought to market. A prerequisite for obtaining valuable patents is to investigate whether or not a new invention is patentable or not, i.e. to check its patentability before the patent application is going to be filed and examined by patent offices.

Moreover, to expand freedom to operate and avoid infringement, it is often necessary to invalidate competitor's patents that might hinder a company's own business. To successfully oppose against third party patents it is essential to find valuable arguments against patentability of the competitor's invention, and to prove that the patent has been granted erroneously because relevant prior art has not been identified during the examination process of the patent office.

For investigating patentability of new inventions and invalidation of competitor's patents it is crucial to perform prior art (or state of the art) searches aiming to analyze if an invention fulfils the main requirements for receiving a patent as specified by patent law, and to avoid double inventions (i.e. novelty/patentability searches) or to find arguments against validity of competitor's patents (i.e. validity searches). The main requirements for patentability defined by patent law are novelty (i.e. claimed subject matter is sufficiently different from prior art, and was not disclosed elsewhere before the filing date of the patent application), and inventive step or non-obviousness (i.e. invention, having regard to the state of the art, must not be obvious to a person skilled in the art). Definitions of prior art given by most systems of patent law from different countries and authorities are very similar. As an example, Article 54 of the European Patent Convention (EPC) may be mentioned [1], which states that

> *"The state of the art shall be held to comprise everything made available to the public by means of a written or oral description, by use, or in any other way, before the date of filing of the European patent application"*.

As a consequence, it is irrelevant if prior art is disclosed in patent literature or alternatively in non-patent literature, including scientific/technical journals and all other kind of content available e.g. on the internet and elsewhere. For in depth analysis of prior art comparable search efforts should therefore be undertaken for non-patent literature as for patent literature.

2. Finding the needle in the haystack: why full-text retrieval of non-patent literature is crucial for prior art searching

Unlike Freedom to operate or infringement searches, which may be limited to the claims of patent literature, prior art searches

☆ Some of the subject matter of this article is based on presentations of the author at the International Patent Information Conference IPI-ConfEx in March 2008 in Seville, Spain, and the International Chemical Information Conference ICIC in October 2008 in Nice, France.

* Tel.: +41 61 3237824; fax: +41 61 3237559.
E-mail address: maik.annies@syngenta.com

0172-2190/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wpi.2009.05.006

should take into account the full-text of a publication, since relevant subject matter does not necessarily need to be in focus, but instead hidden in the full-text. This hidden prior art is often insufficiently or not at all indexed and abstracted in bibliographic databases, and therefore difficult to retrieve, since indexing often focuses on major aspects of the publication, while hidden information is sometimes overlooked or not considered as being relevant. Moreover, due to the diverse indexing policies of bibliographic databases often focussing on certain aspects or parts of the publication, and excluding specific information disclosed in the full-text, important prior art might be missed. As a consequence, for prior art searching it is absolutely crucial to enhance retrieval from bibliographic databases by full-text searching of patent- and non-patent literature.

3. Sources of full-text non-patent literature

Compared to patent literature, which is retrievable from a limited number of patent databases, non-patent literature, such as scientific publications in journals, meeting reports or abstracts of academic theses, is scattered across many different sources. Overall, this makes searching and retrieval of non-patent literature more complicated than patent literature searching. One major source of non-patent literature for professional searching is the various databases supplied by classical online-hosts, including pure non-patent literature databases as well as databases allowing retrieval from both non-patent- and patent literature. However, while many full-text databases for patent literature are available, most non-patent literature databases of classical online-hosts are bibliographic databases, while hardly any full-text database comprehensively covering journal articles from different publishers is offered. One of the main reasons is most likely the copyright-protection of full-text journal articles by publishers, whose commercial interest complicates multi-publisher full-text databases. This raises the question if alternative sources for full-text non-patent literature prior art are currently available, and in how far they are applicable to enhance retrieval from commercial bibliographic non-patent literature databases in professional prior art searching.

One alternative source for full-text non-patent literature is publisher's or host's websites, nowadays offering most featured journals in electronic form as e-journals. These websites often also provide search engines, allowing searching, retrieval and display of current e-journal article content and backfiles.

Finally, the internet with its enormous amount of searchable scientific and technical information is another major source for both patent- and non-patent literature content. However, since this information is often insufficiently structured and edited for searching compared to professional databases, specific tools supporting internet prior art retrieval are essential for professional searching. This becomes even more obvious, when non-computer-readable formats, such as chemical structures published on the web, need to be analyzed.

4. Prior art searching in e-journals

Articles published in scientific and technical journals in electronic form as e-journals, can be displayed and downloaded by the user in different formats such as HTML or PDF. Moreover, publishers and hosts of e-journals provide various tools for searching and retrieval of their e-journal content on their websites. Since often simultaneous searching across various e-journals is offered, for satisfactory retrieval only a limited number of search sites need to be searched. However, unfortunately many search sites are not useful for professional prior art searching, since they lack advanced search and display features critical for fast and effective retrieval and evaluation of e-journal content.

4.1. Minimum requirements for full-text searching

To identify appropriate search sites, three minimum requirements were defined:

(1) Search sites should cover a larger number of e-journals, ideally from as many different publishers as possible and covering technologies of relevance.
(2) Search sites should provide advanced search options, e.g. at least Boolean logic or wildcards should be supported.
(3) Search sites should support advanced display features, e.g. at least the searched keywords should be highlighted within the context of the full-text, since otherwise analysis of retrieved full-text documents tends to be too complicated and time-consuming.

Based on these minimum requirements e-journal search sites for about 2000 different e-journals subscribed to by Syngenta mainly dealing with chemistry and plant biotech were analyzed. From all search sites, only four finally fulfilled the minimum requirements defined above (see Fig. 1) [2–5]. These were reviewed in more detail with regards to their search and display features.

4.2. Analysis of search features

Search screens of the four publisher's or host's e-journal full-text search sites fulfilling the minimum requirements regarding search and display defined above are basically similar. As a representative example Fig. 2 shows the search screen of ScienceDirect, mainly covering scientific journals published by Elsevier Science. In the advanced mode relevant search terms may be entered into two input boxes. Boolean operators may be chosen from pull-down menus or directly entered within the input boxes. Search terms can be searched in full-text or alternatively in other search fields, such as title or abstracts. In addition to scientific journals, it is also possible to include a selection of books to the search. Either all featured journals or only a selection, e.g. subscribed journals, may be searched. To narrow down answer sets specific subject areas may be defined, such as certain scientific disciplines. Finally, a time range for the search may be defined. Other search options offered by the different search sites are e.g. phrase searching (normally applied by using double quotation marks), automatic stemming, wildcards, and citation searching.

In contrast to other e-journal search sites, ScienceDirect also features an expert mode, supporting more complex search strategies, which may be entered within a larger input box (see Fig. 3).

	Number of Journals
ACS Journals	38
Annual Reviews	38
Highwire Press	~ 1200
Science Direct (Elsevier)	~ 2500

Fig. 1. E-journals search sites fulfilling the minimum requirements for professional full-text prior art searching with number of searchable journals.