Patent Citation Indexing and the Notions of Novelty, Similarity, and Relevance*

EUGENE GARFIELD,
Institute for Scientific Information, 325 Chestnut Street, Philadelphia, Pa. 19106

Received September 15, 1965

The unique features of the "references cited" in U. S. patents are discussed in relation to their use in the patent section of the *Science Citation Index*, which adds a new dimension to patent searching. Citation indexing provides a new basis for clarifying the concepts of similarity, coupling, novelty, and relevance.

The original title of this paper illustrates the inherent ambiguity of natural language. "What is relevant in a patent search?" may refer to the types of documents included in the term "prior art." Many types of published documents are "relevant"—patents, journal articles, books, etc. Whether or not the subject matter of a particular document is relevant, is another question. Furthermore, the specific purpose of a patent search affects relevance. Frequently, only the searcher can determine relevance (*1*). Those who have filed patent applications know too painfully how the inventor and the examiner can disagree on what is relevant. What is relevant to one man may be irrelevant to another. There is no objective measure of relevance.

On the other hand, similarity (*2*) is an objective relationship that exists between two documents. Similarity can be measured in several ways. These are not yet precise measures. They are relative. One measure of similarity is word or descriptor coupling (*3*), another is bibliographic coupling (*4*).

Key words or descriptors are natural language terms used in conventional indexing systems as in the *Uniterm Index to Chemical Patents* or *Chemical Abstracts*. In the Uniterm system, the number of Uniterms shared in common by two patents determines their similarity. If the same set of Uniterms is used to index two different patents, either the patents are essentially the same or the indexing has not been sufficiently deep to reveal their dissimilarity. The same would be true of two patents indexed by *CA*.

Bibliographic coupling is based on citation indexing. In citation indexing, the footnotes or references used by authors in writing technical papers are the indexing terms (*5*). The *Science Citation Index*, including its *Patent Citation Index*, is based on citation indexing. In this system, the similarity between two citing documents is a function of the reference citations they share in common. Theoretically, if two different papers contain the same list of "references," then they are essentially the same. If not, as in word indexing, the number of citations is not sufficient to establish their dissimilarity.

Patents, however, are a special case. In patents, there are two kinds of reference citations: those occasionally provided by the inventor, and those provided more frequently by the patent examiner. It is of sociological interest to note that the examiner is comparable to the referee of a technical paper. It is a proper function of the referee (or editor) to determine if an author has cited pertinent prior art. The inventor affirms that to the best of his knowledge his invention is novel. The law does not require that he search the literature or consult his peers to determine the validity of his declaration. This is left to the patent examiner.

The examiner's prior art search frequently turns up a list of pertinent references. These references are usually the basis for disallowing one or more claims. The list of these "references cited," which sometimes includes the inventor's own references, is published at the end of each patent. In the *Science Citation Index*, all of these "references cited" are included as indexing terms. The *Index* does not include all references by the inventor which appear in the body of the specification. To do so would require the expensive task of reading each patent word by word. The inventor's references could be economically included if such references were published in one prescribed location at the beginning or end of the patent. The inventor includes these references to show the state-of-the-art, to identify a priority application, a co-pending application, a continuation, etc.

The majority of references, however, are provided by the examiner and constitute the prior art which the examiner used to disallow one or more claims. Obviously if all claims are disallowed (and about 50% of all patent applications fall in this category), the patent is not issued. The examiner, by definition, cannot cite pertinent anticipatory prior art for the allowed claims, though in fact many cited references are included which did not result in disallowance.

How relevant are these references to the subject matter of any given search? Obviously the examiner considers them relevant "enough" to disallow claims. Anyone interested in learning his reasons can examine the "wrapper" containing the complete file of correspondence. The high frequency of requests for these "references cited" was the reason for listing them from February 4, 1947

*Revised version of paper presented under title "What is Relevant in a Patent Search?" at the Joint Meeting of Division of Chemical Literature and Division of Chemical Marketing and Economics, 150th National Meeting of the American Chemical Society, Atlantic City, N. J., September 1965.

to the present. For patents issued before 1947, it is still necessary to consult the wrapper.

In the past few months, the Patent Office has taken another important step in helping the searcher. Next to each cited patent, its classification number is also given.

In using the *Patent Citation Index*, one is not as much concerned with the question, "Is the cited patent relevant to the citing patent?" as the converse and more significant question, "Is the citing or the *retrieved* patent relevant to the cited patent?" The cited patent is the starting point of the citation index search. This is frequently forgotten or misunderstood by those who have not used a citation index. Patent attorneys should not have this difficulty as they are used to the citator systems long established in legal searching—*e.g., Shepard's Citations* (*6*).

If one has a particular patent in mind, it may be vital to know whether the technology disclosed has been modified, improved, or utilized in any way. This can be done quickly using the *Patent Citation Index*. Arranged in numerical and thereby chronological order, the cited patent is quickly identified. As shown in Figure 1 (sample page of the *PCI*), after each cited patent there is a list of citing patents and/or journal articles. Most of the citing documents will be patents, and all the citing patents are U. S. Patents. However, cited foreign patents are included.

Having found one or more citing patents, the searcher can now turn to the *SCI Source Index* (Figure 2) which provides the full bibliographic description of the citing patent including all inventors, assignees, patent title, classification number, date of issue, and the number of references cited in the patent. He can then decide whether to examine the patent itself or its "abstract," that is, the principal claim in the *Official Gazette*, or its abstract in *Chemical Abstracts*.

Is the retrieved patent or the citing patent relevant? The answer cannot be categorically black or white. It is always some shade of gray which only the searcher can determine. Consider some specific circumstances. A particular 1949 patent describes subject matter which the searcher has determined is relevant. He looks up the patent in the 1964 and 1965 *Patent Citation Indexes* and finds a few 1964 or 1965 patents which have cited it. The examiner cited the 1949 patent because he considered it anticipatory prior art. For this reason, he disallowed one claim which does *not* appear in the list of allowed and published claims. The original claim could be seen in the wrapper. The subject matter of the specification has not been altered one iota. The crux of the question is this: What is the degree of similarity, in any given patent, between the specification and the ungranted claims? It has rarely been my experience to find patents containing completely dissimilar claims. They may be specific embodiments of applications of a general method—*e.g.*, two different specific chemical compounds or two different generic substituents. But even if we found a patent that had two completely dissimilar subject matters claimed, the information disclosed in the specification is the main question.

In a patent once issued to me on a selective copying device, the examiner cited a seismographic recording device. Would the searcher interested in the seismographic recording device patent consider my patent relevant? The writing unit of a selective copying device is a recording instrument! It is not possible to determine relevance on an *a priori* basis. One can predict a given degree of similar-

```
1 163 220--------*15*COREY 112/140-------US
      TIMM RG          3192885 US      P 65
1 163 224--------*15*DODGE 34/124--------US
      PENDLETO.PL      3208513 US      P 65
1 163 251--------*15*MILLSPAUGH 162/217X-US
      FLEISSNE.H       3197896 US      P 65
1 163 296--------*15*WILLIAMS 66/111-----US
      WIGNALL N        3222891 US      P 65
1 163 335--------*64*WEISS G-------------GER
      PAETZOLD R       Z ANORG A C     65  338   22
1 163 339--------*15*HAUSS---------------US
      SCHNEIDE.KJ      3191662 US      P 65
1 163 349--------*15*KIMBALL 40/130------US
      GURIAN SD        3212080 US      P 65
1 163 350--------*15*LEWIS 24/160--------US
      PINTAREL.R       3208239 US      P 65
1 163 358--------*58*  -------------------FRAN
      LOWRIE HS        3170929 US      P 65
      LOWRIE HS        3170930 US      P 65
1 163 375--------*15*SELFRIDGE-----------US
      GREBNER F        3178856 US      P 65
      WENTHE RG        3216875 US      P 65
1 163 401--------*58*  -------------------FRAN
      HANEL E          3192741 US      P 65
1 163 402--------*15*GILLIN--------------US
      SCHLISIN.AE      3167601 US      P 65
1 163 413--------*58*  -------------------FRAN
      PHILLIPS RS      3200406 US      P 65
1 163 448--------*15*PENKALA 25/17--------US
      RAAB HA          3212432 US      P 65
1 163 465--------*64*  -------------------DAS
                       ATOMPRAXIS      A 65  11  295
1 163 475--------*58*  -------------------FRAN
      SEDDON JW        3216202 US      P 65
1 163 513--------*58*  -------------------FRAN
      KLASS DL         3221849 US      P 65
1 163 525--------*58*  -------------------FRAN
      BROWN EA         3199650 US      P 65
      GIANELLI L       3172438 US      P 65
1 163 539--------*64*  -------------------GERM
      MEALS RN         ANN NY ACAD     65  125  137
1 163 541--------*15*HULTIN--------------US
      WAKEMAN AH       3182971 US      P 65
```

Figure 1. Section of *Patent Citation Index* as reprinted from 1965 *Science Citation Index*.

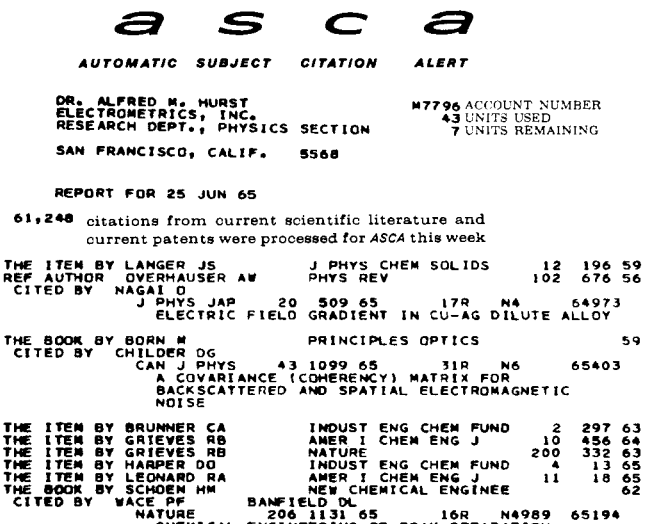
German Patent No. 1,163,335, issued to G. Weiss in 1964, has been cited by R. Paetzold in *Zeit. Anorg. Ang. Chem.* **22**, 338 (1965). The full title of the citing article appears in the *Source Index*. U. S. Patent No. 1,613,375, issued to Selfridge in 1915, has been cited by two 1965 patents issued to Grebner and to Wenthe.

```
PAETZOLD R    AMOULONG H
  Z ANORG A C  337  225 65      24R   N5/6   66190
    UNTERSUCHUNGEN AN SELEN-SAUERSTOFF-
    VERBINDUNGEN .28. DIE SCHWINGUNGSSPEKTREN VON
    KRISTALLISIERTEM FLUSSIGEM UND GASFORMIGEM
    SELENTRIOXID
PAETZOLD R    RONSCH E
  Z ANORG A C  338   22 65      14R   N1/2   67121
    DIALKYLAMIDODERIVATE DER SELENIGEN SAURE
PAETZOLD R    RONSCH E
  Z ANORG A C  338  195 65      10R   N3/4   67752
    UNTERSUCHUNGEN AN SELEN-SAUERSTOFF-
    VERBINDUNGEN .31. ALKANSELENINSAUREALKYLESTER
PAETZOLD R    AMOULONG H
  Z CHEM        5  435 65    M ND R   N11    70270
    KONSTITUTION DER WASSERFREIEN SELENSAURE UND
    VON H2SE04/SE03-LOSUNGEN


GREBNER F     SPINDLER G
  3173193 US              65   P   4R   MAR 16
    CL29/155 METHOD OF MANUFACTURING A LATTICE
    GIRDER
GREBNER F
  3178856 US              65   P  14R   APR 20
    CL50/293 FORM ARRANGEMENT FOR CASTING
    TRANSVERSE CONNECTING MEMBERS IN A CONCRETE
    GIRDER CEILING
GREBNER F     KOELSCH W     LAUTERBA.W    RHEINB GMBH
  3198219 US              65   P   7R   AUG  3
    CL140/112 APPARATUS FOR PRODUCING A GIRDER


WENTHE RG     JOHN L SEYM
  3216875 US              65   P  11R   NOV  9
    CL156/154 METHOD OF MAKING STRUCTURAL
    MATERIALS
```

Figure 2. Source Index showing full titles, subclasses, etc. for citing patents or articles.

ity between the citing and cited patents by examining the primary classifications to which each was assigned. These classifications are included in the *Patent Citation Index* because they are provided, as mentioned earlier, in the published patents.

The provision of the classification numbers in the title of the citing and in the cited patents provides useful information during a *Citation Index* search, but it is perhaps even more helpful in the *ASCA* system. In this current alerting system, the subscriber receives a weekly report informing him where any given patent has been cited in current journal articles or U. S. patents. He can also be notified of all currently issued patents which fall into a particular classification or those assigned to a particular company. He can also use an inventor's name as part of his interest profile or any specific technical paper or book ever published. The scope of this service is quite large, involving at present about 1500 leading journals and all U. S. patents—over 3,000,000 reference citations per year appearing in 235,000 source papers and 60,000 U. S. Patents in 1965. A copy of a typical *ASCA* report is shown in Figure 3.



a s c a

AUTOMATIC SUBJECT CITATION ALERT

DR. ALFRED M. HURST
ELECTROMETRICS, INC.
RESEARCH DEPT., PHYSICS SECTION
SAN FRANCISCO, CALIF. 5568

M7796 ACCOUNT NUMBER
43 UNITS USED
7 UNITS REMAINING

REPORT FOR 25 JUN 65

61,248 citations from current scientific literature and current patents were processed for *ASCA* this week

THE ITEM BY LANGER JS — J PHYS CHEM SOLIDS — 12 196 59
REF AUTHOR OVERHAUSER AW — PHYS REV — 102 676 56
CITED BY NAGAI O
J PHYS JAP 20 509 65 17R N4 64973
ELECTRIC FIELD GRADIENT IN CU-AG DILUTE ALLOY

THE BOOK BY BORN M — PRINCIPLES OPTICS — 59
CITED BY CHILDER DG
CAN J PHYS 43 1099 65 31R N6 65403
A COVARIANCE (COHERENCY) MATRIX FOR BACKSCATTERED AND SPATIAL ELECTROMAGNETIC NOISE

THE ITEM BY BRUNNER CA — INDUST ENG CHEM FUND — 2 297 63
THE ITEM BY GRIEVES RB — AMER I CHEM ENG J — 10 456 64
THE ITEM BY GRIEVES RB — NATURE — 200 332 63
THE ITEM BY HARPER DO — INDUST ENG CHEM FUND — 4 13 65
THE ITEM BY LEONARD RA — AMER I CHEM ENG J — 11 18 65
THE BOOK BY SCHOEN HM — NEW CHEMICAL ENGINEE — 62
CITED BY WACE PF BANFIELD DL
NATURE 206 1131 65 16R N4989 65194
CHEMICAL ENGINEERING OF FOAM SEPARATION

REF AUTHOR WEINER — FEB INT SOL STAT CIR — 32 62
CITED BY WEINER PK RAD CORP AM
3191061 US 65 P 7R JUN 22
CL307/88.5 INSULATED GATE FIELD EFFECT DEVICES AND ELECTRICAL CIRCUITS EMPLOYING SUCH DEVICES

THE PATENT 1900018 LILLIENFELD 317/258 US 33
CITED BY BOYKIN OF CTS CORP
3191108 US 65 P 11R JUN 22
CL317/258 ELECTRICAL CAPACITOR AND METHOD OF MAKING SAME

Figure 3. Typical weekly ASCA report illustrating citation of a 1933 U. S. Patent by a 1965 Patent. Other profile items include journal articles, books, authors, etc.

It has been our experience that users of this system have found a high degree of *pertinence* in the patents and papers disseminated by the *ASCA* service or retrieved by the *Science Citation Index*. Since there is no objective measure of relevance, we would suggest that critics evaluate the system of citation indexing on the basis of *a posteriori* user judgements rather than any *a priori* and ill-conceived notions of relevance.

It has been said that a citator system is necessary and useful for the lawyer because American law is based on the "doctrine of *Stare Decisis* which means that all courts must follow precedents laid down by higher courts and each court generally also follows its own precedents" (*6*). This has been misconstrued as the raison d'etre for the citator system. On the contrary, it is because the lawyer "must make sure that his authorities are still good law, that is, that the case has not been overruled, reversed, limited or distinguished in some way that makes it no longer useful as a valid authority. Here is where the use of *Shepard's Citations* comes in.... The amazing efficiency of the citation method is such that once the starting case or statute is found it becomes a key that unlocks the entire store of law on a given point" (*6*).

By analogy, patent searches involve not only what is commonly called "prior at" but also what may be called "subsequent art." Technological innovations are not conceived in a vacuum; nor are they pulled from the air by magic. Every patent involves one or more primordial concepts which the inventor has joined together in a unique way to justify his claim for patent protection. Finding the needles in the haystack—the pertinent patents or publications—rapidly and efficiently, is the function of any index. The *Patent Citation Index*, if properly used in combination with existing tools, can save many valuable hours of search time and also make the time spent in searching productive of information that would otherwise be difficult or impossible to uncover.

ACKNOWLEDGMENT

I should like to acknowledge the help of R. A. Spencer of the U. S. Patent Office (*7*) in clarifying many points discussed in this paper.

LITERATURE CITED

(1) Taylor, R. S., *Am. Doc.* **13**(4), 31-96 (1962).
(2) Salton, G., *J. Assoc. Computing Machinery* **10**(4), 440-57, (1963).
(3) Garfield, E., "World Brain or Memex?—Mechanical and Intellectual Requirements of Universal Bibliographicl Control." Paper presented at the Symposium on "The Foundations of Access to Knowledge," Syracuse University, Syracuse, N. Y., July 30, 1965, in press.
(4) Kessler, M. M., *Am. Doc.* **14**, 10-25 (1963).
(5) Garfield, E., *Science* **144**(3619), 649-54 (1964).
(6) Adair, W. C., *Am. Doc.* **63**, 31-32 (1955).
(7) Spencer, R. A., private communication to E. Glazer, May 13, 1964.

See also:

(8) Garfield, E., *J. Pat. Off. Soc.* **39**(8), 583-95 (1957).
(9) Seidel, A. H., *ibid.*, **31**, 554 (1949).
(10) Hart, H. C., *ibid.*, **37**, 714 (1949).
(11) Sokal, R. R., and Sneath, P. H. A., "Principles of Numerical Taxonomy," San Francisco, Calif., W. H. Freeman and Co., 1963.

NOTE: Since the presentation of this paper, the *Science Citation Index* service has eliminated the indexing of U. S. Patents as sources. However, the *SCI* continues to index all patents cited in approximately 1500 technical journals as typified by the Paetzold article in Figures 1 and 2.