

Garfield, E. "What Is Relevant in a Patent Search?" p.1-11, September 1965. Presented at the American Chemical Society, joint meeting of divisions of Chemical Literature and Chemical Marketing and Economics; Atlantic City, NJ.

What is Relevant in a Patent Search? \*

by

Eugene Garfield, Ph.D., Director  
Institute for Scientific Information  
325 Chestnut Street  
Philadelphia, Pa. 19106

The 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.

\* Presented at the American Chemical Society, Joint Meeting of Division of Chemical Literature and Division of Chemical Marketing and Economics; Atlantic City, N. J.; September, 1965.

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 usually turns up a list of pertinent references. These references are frequently 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, we include all of those references as indexing terms. We do not process as yet, those occasional references appearing in the specification. These would require the expensive task of reading each patent but could be economically included if such references were published in one prescribed location, such as is done for the "references cited," at the end of the patent.

The majority of references, therefore, are provided by the examiner and constitute the prior art which the examiner used to disallow one or more claims.

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 usually turns up a list of pertinent references. These references are frequently 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, we include all of those references as indexing terms. We do not process as yet, those occasional references appearing in the specification. These would require the expensive task of reading each patent but could be economically included if such references were published in one prescribed location, such as is done for the "references cited," at the end of the patent.

The majority of references, therefore, 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 obviously, by definition, cannot cite 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. The high frequency of requests for these "references cited" was the reason for listing them from February 4, 1947, 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, we are 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 as, 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 presently 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 which provides the full bibliographic description of the citing patent including all inventors, assignees, patent title, classification number, date of issuance, and the number of references cited in the patent. He can then decide whether to examine the patent, or its "abstract" in the Patent Gazette or Chemical Abstracts. Patents are processed and available to users of our Automatic Subject Citation Alert (ASCA) system within an average of three weeks. However, the SCI is issued quarterly and cumulated annually. The average time lag in the printed Indexes is about three months.

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 or applications of a general method, as, 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 presume a given degree of similarity between the citing and cited patents by examining the primary classification to which each was assigned. These classifications are now 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 1,100 leading journals and all U. S. patents--some 3,000,000 reference citations per year appearing in 225,000 source papers and probably 75,000 U. S. patents in 1965 at the present rate of issuance. A copy of a typical ASCA report is shown in Figure 2.

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 prefer that our critics evaluate the system on the basis of a posteriori user judgments rather than any a priori and ill-conceived notions of relevance.

In closing I should like to refer to some correspondence between the Institute for Scientific Information and the U. S. Patent Office Research Department on the problem discussed in this paper. In order not to avoid any possible misinterpretation, I will quote verbatim the comments of R. Spencer (7) and my reply (8).

Mr. Spencer wrote:

"It is possible, although not too likely, that the examiner might refer to some other reference in the body of the letter. Such a citation would not be listed at the end of the patent. It is not unusual for the applicant or attorney to cite related art in his letters. Such citations are not listed unless the examiner specifically cites them (in the proper form) in a subsequent office action. In addition references are cited in the body of the specification. Such citations are included to show the state of the art, to identify a priority application, to identify a copending application or to identify an application of which the instant one may be a division, a continuation in part or a continuation, etc. Such citations, although they may be very closely related to the subject matter of the patent, are not listed. Although application numbers are generally given in cases as originally filed, the corresponding patent numbers are frequently added by amendment if such applications mature into patents during the pendency of the instant one.

"In almost all case, references are 'cited' for their substantive content. The citation might be related to a legal issue such as a requirement for restriction or division or a rejection for double patenting etc., but it is the substantive content of the document that supports the legal requirement. However, there need not be a close relationship between the content of the cited document and the content of the application.

"I think the real point is that the use of citation indexing is based on the theory that cited references have a high probability of being related to the document in which they are cited. In the case of patents the cited references not listed at the end may be the most closely related. The examiner may not have been able to use such references during the pendency of any particular patent but he might very well be able to use them to apply to some other application. Further, the listed references are those the examiner selected as being the closest prior art with respect to the claims that were presented throughout the prosecution. Therefore, all the listed references are not necessarily related to disclosed but not claimed subject matter or to the allowed claims."

My reply stated:

".....most references listed at the end of a patent are due to their substantive, and not legal, content. .... he will be directed to subsequent documents, patents, and journal articles, .... The degree of relevance between the cited and the citing patents is always relative. There is no measure of relevance, of which I am aware, that anyone can use at present for evaluating any system, including the SCI."

".....the 'references cited' are far from exhaustive of the documents which could be cited in a patent. Furthermore, our present processing only includes such 'references cited' and not the references appearing in the text of the patent specification. This is a step we can take at a time when we can justify the expense of the 'pre-edit' required to extract these extra citations."



".....While it may or may not be true....that the cited references not listed at the end of a patent are more closely related to the subject matter of the source (citing patent), this in no way affects the high degree of relevance between those that are listed and the citing patent. Stated another way, the citing patent might not be retrieved on a search beginning with the unlisted.....patent. The citing patent would be retrieved on a search beginning with those that are listed. This dichotomy is also true of the existing classification system. Patents are classified on the basis of claimed subject matter and not on the basis of subject matter disclosed or referred to in the specification.

"As to whether the unlisted cited references are 'the most closely related,' the question is: Related to what? .....the assumption that one 'descriptor' (cited reference) is more relevant than another 'descriptor' shows that you are thinking in terms of the conventional retrieval approach. If Spencer means that the most novel aspect of an invention, and for which the inventor is granted a patent, is not 'covered' by any of the references cited, ..... there probably does not exist any pertinent literature or patent for the inventor or the examiner to cite with regard to that novel aspect. The citations which the inventor can and may provide as background, e.g., to show what has gone before, will be unmeasurably relevant to his invention. So will the citations which were used to reject some of his claims. As to which would be most closely related, I hesitate to predict as it obviously will vary. ....

"If there existed in the literature a perfect match between two patents, then there would have been, ipso facto, no novelty. A citing patent will be retrieved by SCI due to 'closeness' of the allowed claims and the rejected claims in a single invention. If enough prior art exists, the citations pile up until there is no novelty or patentable invention at all. I agree there is a real possibility that the disclosed but unclaimed subject matter of a patent specification may contain information not disclosed anywhere else. However, I don't know whether this information is retrieved any more effectively through the unlisted or listed references. I could imagine cases of both kinds. ....

"We would welcome the addition of the presently unlisted references in the form of a bibliography at the end of the patent specification. This would improve the overall effectiveness of the SCI with a reasonable expenditure of energy. However, it is not necessary for the effectiveness of the SCI system to have a complete listing of every possible document one could cite. This is the very same principle we have found in handling the journal literature. An author is selective in the citations he chooses. This is in part conditioned by how well the prior art is known.

"It is not necessary for every telephone improvement patent to cite Alexander Graham Bell. If someone filed on Bell's invention tomorrow, we would expect that the examiner would be knowledgeable enough to cite the Bell patent. If the new application had one particular novel feature and a claim were granted, we would consider it pertinent for the examiner to retrieve that improvement patent each time he looked up the Bell citation in the SCI. In subsequent improvement patents, the earlier improvement patent would tend to be cited. ....one operates on the assumption that the examiner or scientist has some knowledge of



"the prior art and can begin his search from there. That the first dozen claims in the above-mentioned fictional case would be rejected by reference to the Bell patent would not diminish its relevance to a search on telephone art. Nor is that citation less relevant than a citation the inventor might make to the Morse telegraph..... The relevance would be a function of the particular search. In a search on telephone art, it would be relevant. And on a search of telegraph art, it might not be relevant. If the invention were admittedly an improvement on the Bell telephone, then the inventor would cite the Bell patent, and the examiner might not be able to find any other prior art. In that case, it would be a shortcoming of the SCI not to include the inventor's reference."

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, the patent search involves not only what is commonly called "prior art" 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. We believe that the Patent Citation Index does this and, if properly used in combination with existing tools, cannot only save many valuable hours of search time but also make the time spent in searching productive of information that would otherwise be difficult or impossible to uncover.

### References

1. R. S. Taylor, "The Process of Asking Questions," American Documentation 13(4), 391-396 (1962).
2. G. Salton, "Associative Document Retrieval Techniques Using Bibliographic Information," Journal of the Association for Computing Machinery 10(4), 440-457 (1963).
3. E. Garfield, "World Brain or Memex?--Mechanical and Intellectual Requirements for Universal Bibliographic Control." Paper presented at the Symposium on "The Foundations of Access to Knowledge," Syracuse University, Syracuse, N. Y., July 30, 1965. (To be published).
4. M. M. Kessler, "Bibliographic Coupling between Scientific Papers," American Documentation 14, 10-25(1963).
5. E. Garfield, "Science Citation Index--A New Dimension in Indexing," Science 144(3619), 649-654 (1964).
6. W. C. Adair, "Citation Indexes for Scientific Literature?" American Documentation 6, 31-32 (1955).
7. R. A. Spencer, Memorandum to E. Glazer, dated May 13, 1964.
8. E. Garfield, Private Communication to E. Glazer, dated August 18, 1964.

See also:

E. Garfield, "Breaking the Subject Index Barrier--A Citation Index for Chemical Patents," Journal of the Patent Office Society 39(8), 583-595 (1957).

A. H. Seidel, "Citation System for Patent Office," (Letter-to-the-Editor), Journal of the Patent Office Society 31, 554 (1949).

H. C. Hart, "Re: 'Citation System for Patent Office,'" (Letter-to-the-Editor), Journal of the Patent Office Society 31, 714 (1949).

FIGURE 1

1965 PATENT CITATION INDEX  
 2nd Quarterly Issue--April to June  
 Reprinted from  
SCIENCE CITATION INDEX

© Copyright 1965 by the  
 Institute for Scientific Information  
 Philadelphia, Pa. 19106

1,161,857

1 161	857	GLENN LL	3180334 US	P 65		
					US	
1 161	888	JOHNSON RE	3183947 US	P 65		
					US	
1 161	895	LYMAN F	3189672 US	P 65		
					GERM	
1 161	906	LEHMKUHL H	ANN NY ACAD	R 65	125	124
					US	
1 162	005	FRISBY VC	3183134 US	P 65		
					DAS	
1 162	052	ATOMPRACTIS		A 65	11	231
					US	
1 162	125	BERRYER P	3190072 US	P 65		
					US	
1 162	246	SCRAGG F	3183730 US	P 65		
					US	
1 162	270	RONNE RM	3189952 US	P 65		
					FRAN	
1 162	339	KUMPF H	3182001 US	P 65		
					US	
1 162	382	RICKERT RJ	3180801 US	P 65		
					US	
1 162	405	GWYN CB	3191272 US	P 65		
					US	
1 162	559	BEILER AP	3182782 US	P 65		
					FRAN	
1 162	572	BETTS WM	3176791 US	P 65		
					US	
1 162	595	HAYGEMAN RM	3189250 US	P 65		
					US	
1 162	607	ANTON M	3184922 US	P 65		
					US	
1 162	698	PETRIN F	3184217 US	P 65		
					US	
1 162	775	BRACH I	3181716 US	P 65		
					FRAN	
1 163	100	TURGEON JA	3192333 US	P 65		
					US	
1 163	105	QUILLIAN RD	3177861 US	P 65		
					FRAN	
1 163	339	BOCKLY E	3177078 US	P 65		
					FRAN	
1 163	375	BUHRENDO.FG	3180549 US	P 65		
					US	
1 163	539	SCHNEIDE.KJ	3191662 US	P 65		
					US	
1 163	586	GREBNER F	3178856 US	P 65		
					GERM	
1 163	589	MEALS RN	ANN NY ACAD	65	125	137
					US	
1 163	590	WAKEMAN AH	3182971 US	P 65		
					US	
1 163	619	VANDERLE.C	3191689 US	P 65		
					US	
1 163	688	BERRILL WH	3182354 US	P 65		
					US	
1 163	734	HARKER JK	3191025 US	P 65		
					FRAN	
1 163	734	VONLOEWI.A	3178167 US	P 65		
					US	
1 163	734	GEYER HM	3177982 US	P 65		
					US	
1 163	734	BAROSKO JM	3182867 US	P 65		
					US	

FIGURE 2



**AUTOMATIC SUBJECT CITATION ALERT**

a service of the **INSTITUTE FOR SCIENTIFIC INFORMATION**

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 MICHAEL E T AMER MATH SOC 71 152 51  
CITED BY SMITHSON RE  
P AM MATH S 16 448 65 6R N3 66090  
CHANGES OF TOPOLOGY AND FIXED POINTS FOR  
MULTI-VALUED FUNCTIONS

THE ITEM BY GOLDSTEIN R NUCL SCI ENG 13 132 62  
CITED BY KELBER CN  
NUCL SCI EN 22 120 65 L SR N1 64646  
A SIMPLE ESTIMATE OF EFFECTS OF RESONANCE  
INTERFERENCE

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 ENGINEER 62  
CITED BY WACE PF BANFIELD DL  
NATURE 206 1131 65 16R N4989 65194  
CHEMICAL ENGINEERING OF FOAM SEPARATION

REF AUTHOR WEIMER FEB INT SOL STAT CIR 32 62  
CITED BY WEIMER BY 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

ACCT NO M7796