Polymers and Patents Don't Mix—Easily[†]

MARY ZURBACH BALENT* and JOHN W. LOTZ

IFI/Plenum Data Company,[‡] Wilmington, Delaware 19808

Received January 23, 1979

Polymer references in the patent literature pose a unique challenge to the indexing-retrieval process. Of the several forces working together to compound the problem, four appear to be significant: (1) the wide range in specificity and format of the references, (2) the wide range in inquirer requirements, (3) the sheer volume, and (4) the indexing cost/search cost relationship. This paper addresses itself to the defining and sorting out of these forces and then to ways of coping with them. Statistics are presented showing the scope and magnitude of the problem. A polymer indexing/retrieval system designed to meet the challenge and including an optimizing combination of generic, specific, and monomer/role indexing techniques to produce a favorable indexing cost/search cost relationship is discussed.

In the book "Understanding Chemical Patents", Maynard stated that "new findings of commercial value are often reported in patents well before they appear in the journal literature".[1] As examples of new areas of chemistry which were first disclosed through patents Maynard cited the development of polyurethane technology and the polymerization of olefins with coordination complex catalysts. At this point it seems significant to note that both examples concern polymer chemistry.

To say that "polymers and patents don't mix—easily" may be an understatement. Patents require a specialized language. Information contained in the patent must be couched in terms that will provide the best protection for the inventor. To sort out the technical meat of a patent from its legal shell can cause problems for the information scientist. Because of his technical background he frequently finds it difficult to accept the ever broadening nature of the patent disclosure. Once he has overcome this first obstacle he faces still another.

In any of the technical literature—patents being no exception—polymers may be described in several ways. A particular polymer may be described in terms of its starting monomers, it may be drawn to show the repeating unit of the polymer with no mention of starting materials, or it may be named specifically. This wide variety of formats for naming or identifying any particular polymer is a second source of frustration to the information scientist. His ordered way of thinking inherently fights this unruly proliferation of options.

Because patents are what they are and because polymers are what they are, the combination poses a dual challenge to the indexing-retrieval process. More specifically, we feel that there are four interrelated forces which, when combined in the patent/polymer environment, compound the indexing-retrieval problem. They are (1) the wide range in specificity as well as in format of the polymer references, (2) the wide range in inquirer requirements, (3) the sheer volume, and (4) the indexing cost/search cost relationship. This paper will define and examine these forces in relation to polymers and patents. An indexing-retrieval system designed to meet the dual challenge and including an optimizing combination of generic, specific, and monomer/role indexing techniques will be discussed.

METHOD

All statistics cited in this paper were generated from information contained in the "IFI Data Base of U.S. Patents—1950 through 1977" maintained on magnetic tape by IFI/Plenum Data Co. at Arlington, Va. This file contains:

Patent Titles
U.S. Patent Office Classification
Assignee Names
Inventor Names
IFI Keyword Indexing

IFI Keyword indexing is done from the full text patent using systematic procedures and three controlled open-ended vocabularies—General Term, Compound Term, and Fragment Term. The General Term vocabulary includes references to natural materials, tradenames, reactions, uses, polymer class terms, and nonstructurable chemicals. The Compound Term vocabulary includes references to over 12 000 specific compounds. The Fragment Term vocabulary is used to describe compounds in terms of substructural pieces that characterize them.

The IFI polymer system which will be described in detail later in this paper is part of the IFI Keyword indexing system and uses all three vocabularies as well as roles to identify polymers.

The IFI file was processed using conventional computer techniques to provide the data necessary for summaries and graphs noted in the following sections.

POLYMER INDEXING/RETRIEVAL FORCES

Specificity and Format. Polymer references in the patent literature may range anywhere on a continuum from the very broad or generic to the very narrow or specific. Patents may refer very broadly to a polymer, or to an acrylic polymer, perhaps an acrylic copolymer, or to a vinyl ester-acrylic copolymer, or even very specifically to a vinyl acetate-acrylic acid copolymer.

For every point on the specificity continuum there is another point on a format continuum with which we must also contend. As mentioned previously polymers may be described in any of several ways. Thus the specific polymer polyethylene terephthalate could be named, or it could be drawn to show the repeating polymer unit, or it could be described in terms of its preparation, e.g., a polymer prepared by reacting terephthaloyl chloride and ethylene oxide. A more generic polymer such as polyalkylene terephthalate could be described similarly, replacing the "ethylene" with alkylene. The "terephthalate" portion could be replaced by "a phthalate", by "aromatic dicarboxylate", or simply be mentioned as a dicarboxylate. By using combinations of polymer class words, starting monomers, repeating units, and Markush structures the variations are legion. Somehow the indexing-retrieval process must handle them all.

Inquirer Requirements. One of the most important requirements of a good indexing-retrieval system is that it be responsive to user needs. If a system is not designed to meet

[†] Presented before the Division of Chemical Information, 176th National Meeting of the American Chemical Society, Miami Beach, Fla., Sept 13, 1978.

[‡] A division of Plenum Publishing Co., 227 West 17th St., New York, N.Y. 10011.

0095-2338/79/1619-0080$01.00/0

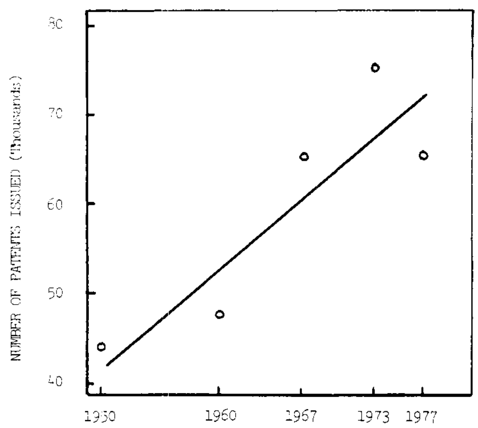

Figure 1. Growth in patent issues, 1950–1977.

Table I. Primary U.S. Patent Office Classes for Polymers.

Class 260	CHEMISTRY, CARBON COMPOUNDS
Subclass 2 thru Subclass 42.57	Synthetic Resins With filler, dye or pigment Polymerized unsaturate With filler dispersion aid
Subclass 45.7 thru Subclass 96	With preservative Processes Subsequent treatment Polymer crystallization
Subclass 823 thru Subclass 901	Mixed synthetic resins, graft copolymers or block copolymer Polymerized ethylenically unsaturated compounds Polymer of an ester of an ethylenically unsaturated monocarboxylic acid
Class 526	SYNTHETIC RESINS, ADDITION POLYMERS (all subclasses)
Class 528	SYNTHETIC RESINS, PROCESSES (all subclasses)

the needs of the group it serves, it will not be effective. Polymer chemists fortunately talk essentially the same language as their counterparts preparing patent specifications. They use the same wide range of specificity levels and identification formats as the patent references themselves. Unfortunately, however, there is no assurance of an automatic one to one match and, because of the variety of acceptable formats, the system must help in this translation step.

Inquiries may range from a specific polymer, e.g., acrylonitrile–vinylidene chloride copolymer, to a search for a generic or family of monomers in a polymer class, e.g., polyesters of dihydroxy aromatic compounds. In some cases the requester is interested in a specific polymer or polymer class as the main inventive feature, but in other cases any mention will suffice. What's relevant in one search may not be relevant in another.

An effective indexing–retrieval system must be flexible enough to meet inquirer requirements with regard to level of query specificity and answer relevance. It should also concern itself with effective means for screening out excessive noise and keeping relevance and recall at practical levels.

Growth in Polymer Patent Information. No one familiar with the patent literature needs to be told that the 1960s and early '70s produced exceptional growth in issue rates. This is graphically illustrated in Figure 1. Fortunately, there has been a gradual tapering off during the last few years. Nevertheless, we, as information users and information suppliers, must still cope with those high issue years while somehow keeping up with the continuing heavy load of current accessions.

Polymer technology was not standing still during those same growth years but expanding at its own accelerated rate. To

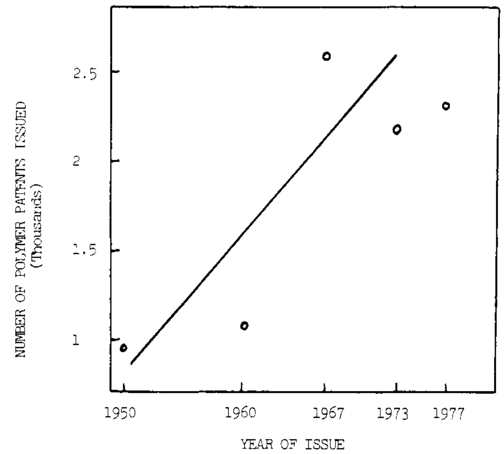

Figure 2. Growth in polymer patent issues, 1950–1977.

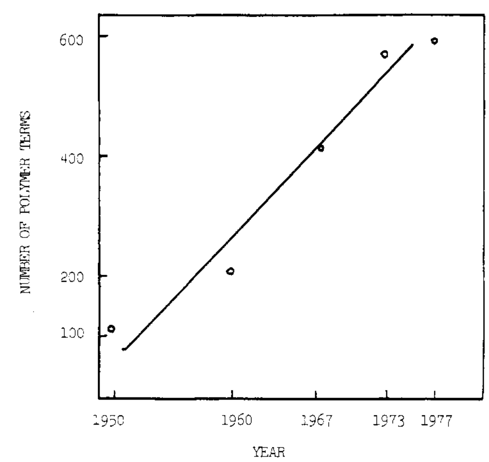

Figure 3. Growth of polymer terms in IFI vocabulary.

illustrate this growth we selected those U.S. Patent Office classes which are devoted primarily to polymers and determined the numbers of patents issued in those classes for the years 1950, 1960, 1967, 1973, and 1977. The classes chosen for this purpose are listed in Table I. In 1950, 790 patents issued in the selected classes. During 1967, 2598 patents issued in these same classes and in 1977, 2361. Figure 2 plots the growth of polymer patent issues from 1950 through 1977. It is important to note that the percentage of polymer patents relative to total issues doubled between these years.

In trying to keep up with this increasing volume, information people have responded in various ways. In 1972, the Patent Office made the unofficial alpha subclasses a recognized part of their system. This increased the number of subclasses in the polymer area from 269 in 1971 to 830 in 1974. The recent addition of the 500 series classes increased the total to 1125.

At the same time IFI was keeping pace through the systematic addition of new polymer terms to the controlled General Term vocabulary. In 1950, IFI used 139 terms to describe polymer concepts. In 1967 the number had grown to 428. In 1971 rights to a DuPont monomer role system were obtained and the system was modified for merger into the existing IFI vocabulary.[2] By 1977 the polymer terms list had grown to 574. This increase in polymer terms is graphically illustrated in Figure 3. The controlled vocabulary list and the role system constitute the salient features of the system now used in the IFI Comprehensive Index.

The sheer volume of polymer references in the patent literature is another factor with which we must be concerned. A look at the number of postings accumulated over a 27-year period by some of the polymer terms in the IFI data base will attest to this statement. The specific polymer term POLY-

ETHYLENE has been indexed for 16292 patents between 1950 and 1970—and people wonder why an information scientist cringes when a customer asks for all information on polyethylene. The term THERMOPLASTIC RESINS shows up in 17051 patents and the class term POLYURETHANES in 12510.

Indexing Costs/Search Costs Relationships. It has been noted that the specificity of disclosures found in polymer patents varies over a wide range, that users have varying needs with respect to search requirements, and that polymer patents are becoming increasingly complex and prolific.

In any discussion of techniques for handling information, it is important also to consider costs, at least in a general way. It is IFI's position that, for any information retrieval system, input or indexing costs must be measured against their relation to and effect on search costs. What one is willing or not willing to spend in screening, i.e., technically evaluating potential answers, must be weighed against what one is willing or not willing to pay in input costs. As others have noted before us, there is a definite relation between input costs and search costs. This relationship may be summarized as follows: the more specifically you index a document the more costly the indexing becomes, and the less costly the search becomes, and conversely, the more generically you index a document the less costly the indexing becomes and the more costly the search becomes.

Of course, the picture is not quite that simple. While precision may increase as you approach the specific end of the indexing scale, it becomes more difficult to maintain indexing consistency. However, assuming for this discussion that this is not a factor and that recall remains the same throughout the specificity range, the point may be made that somewhere along the specificity scale there is an optimum degree of indexing specificity where the input costs have the most favorable cost/value relationship to searching costs.

At IFI we have developed, over the years, a system which we believe is near the middle of the scale—probably a little nearer the specific than the generic end. We feel that it represents and accomodates the interests and needs of a wide range of users. Although we have no hard facts to support it, we feel from experience that it approaches the optimum from the indexing cost/search cost point of view.

IFI POLYMER SYSTEM

The IFI polymer indexing–retrieval system has been designed to meet user requirements at a favorable cost. Questions ranging from the very specific to the generic can be handled by the system with equal ease.

The polymer indexing–retrieval system allows for the identification of most polymers in terms of their corresponding starting monomers.[3] For greater specificity, the more common homo- and copolymers may be indexed as specific items while the uncommon may be indexed merely by a polymer class term. To differentiate between an indexing to a monomeric compound and its corresponding polymer, a unique set of roles is used. This system of roles serves further to indicate the structural class to which the polymer belongs, whether the polymer is a homopolymer or copolymer, whether it is modified or unmodified, and whether the polymer is being physically processed, is being used as a starting material, or is the product of a reaction.

In order to control polymer indexing, and to build greater specificity into the index where needed, polymers are divided into several classes. The main breakdown occurs between what we call the common classes and the "other" classes. By definition, the common classes are those having high frequency of use. They are listed in Table II. The name Polycarbon refers to polymers from unsaturated monomers, having a carbon backbone. The structures of the others are obvious from their names.

Table II. Common Polymer Classes.

Polymer class	Prescribed monomer	Role indicator: Homopolymer	Role indicator: Copolymer
POLYCARBONS	C=C or C≡C	41 51 61	71 81 91
POLYETHERS	HOROH	42 52 62	72 82 92
POLYESTERS	HOROH and	43 53 63	73 83 93
	$HO_2CR'CO_2H$	43 53 63	73 83 93
POLYCARBONATES	HOROH	44 54 64	74 84 94
POLYAMIDES	R"NHR'NHR" and	45 55 65	75 85 95
	$HO_2CR'CO_2H$	45 55 65	75 85 95
POLYURETHANES	HOROH and	46 56 66	76 86 96
	OCNR'NCO	46 56 66	76 86 96
POLYSILOXANES	$HOSi(R)_2OH$	47 57 67	77 87 97

To index polymers from these classes, we have defined arbitrary starting materials or monomers. The arbitrary monomers are shown in Table II and are, in most cases, the obvious ones. The use of arbitrary, or prescribed, monomers eliminates ambiguity at search. If the actual starting materials are important, these are also indexed using monomer reactant roles.

For example, the prescribed monomers for a polyamide are the diamine and the diacid. A patent describes the preparation of a polyamide by reacting an acid chloride with a diamine. If it is significant that the acid chloride was used in the preparation, as it would be in this case, the acid chloride would be indexed as a starting material in the monomer reactant role, while the final product, the polymer, would be indexed in terms of the diacid and diamine in the appropriate polymer roles.

For ease in application, the role numbers have been designed with built-in significance. Two-digit numbers were used. The second digit indicates polymer class, 1 for polycarbons, 2 for polyethers, etc. The first digit shows whether the polymer is being physically processed, is a reactant, or is a product, and also whether it is a homo- or copolymer.

For example, adipic acid, role 63, would be used to index a patent where adipic acid was one segment of a homopolyester, the synthesis of which was described. Adipic acid, role 85, would be used to index a copolyamide which was the reactant in some reaction, e.g., chlorination of the copolyamide. Table II lists the roles which are used to describe the common polymers.

In order to allow for searching polymers by class rather than by specific monomer, each polymer indexing includes reference to the class to which the polymer belongs. For example, all polyesters are not only indexed in terms of their prescribed starting materials, but are also indexed by the term POLYESTERS. For this purpose, the polycarbon class is divided into appropriate subdivisions to provide greater specificity for this very highly posted class. Examples of these terms are:

ACRYLIC HOMOPOLYMERS/ESTER/
VINYL COPOLYMERS/HALIDE/
HOMOPOLYENES/CONJUGATED/

These polymer class terms also take a set of roles which tell whether the polymer is modified or unmodified, and whether the polymer is being physically processed, is a reactant, or is a product. In addition to the polymer class terms, certain very common polymers are also indexed as the specific homo- or co-polymer. For example, in patents describing the preparation of a butadiene/acrylonitrile copolymer the following would be indexed:

ACRYLONITRILE	91
BUTADIENE/1,3-/	91
ACRYLIC COPOLYMERS/NITRILE/	30
COPOLYENES/CONJUGATED/	30
ACRYLONITRILE-BUTADIENE COPOLYMER	30

The monomers in appropriate polymer roles are indexed as well as the polymer class terms and the specific copolymer term ACRYLONITRILE-BUTADIENE COPOLYMER.

Obviously, this procedure allows great flexibility in specificity level when searching. A patent so indexed would be selectively retrieved in a search for the specific copolymer. It would also be easily retrieved in a search for any butadiene or any acrylonitrile polymer or even more broadly any nitrile or any conjugated polyene polymer.

Polymers not included in the seven classes just described are handled somewhat differently. For convenience, these are called Other Polymers. They include (1) polymers containing a repeating backbone structure which is not one of the common types, and (2) polymers having no consistent repeating unit, e.g., polyimides, organometallic polymers. An analogous set of roles is used for these polymers.

The polymer monomers may be described either by the IFI Compound Term vocabulary or by the IFI fragmentation system. Each compound in the Compound Term vocabulary is also described via this fragmentation system so that in a search for polymers prepared from unsaturated isocyanates the requester is able to combine the IFI fragmentation system with the polymer role system for effective retrieval of pertinent information.[4]

CONCLUSION

We believe the polymer indexing-retrieval system we have just described is a viable approach to overcoming some of the obstacles inherent in handling polymer patents: the variety of options for identification, the variety of user requirements and interests, the volume, and the cost relationships. By using combinations of specific polymer names, prescribed starting monomers, roles, and polymer class terms, the searcher is able to handle questions ranging from the very specific to the generic. The system is able to manipulate a large volume of material and is designed to be cost effective from an indexing cost versus search cost point of view. Using this system we feel we are able to make polymers and patents mix more easily.

LITERATURE CITED

(1) J. Maynard, "Understanding Chemical Patents", American Chemical Society Publication, Washington, D.C., 1978, p 23.
(2) L. E. Rasmussen and J. G. van Oot, "Operation of DuPont's Central Patent Index", *J. Chem. Doc.*, **9**, 201-206 (1969).
(3) J. L. Schultz, "Polymer Nomenclature, Classification and Retrieval in the DuPont Central Report Index", *J. Chem. Inf. Comput. Sci.*, **15**, 94-100 (1975).
(4) M. Z. Balent and J. M. Emberger, "A Unique Fragmentation System for Indexing Patent Literature", *J. Chem. Inf. Comput. Sci.*, **15**, 100-104 (1975).

Fossil Fuels in *Chemical Abstracts*†

CHARLES W. MOULTON
Chemical Abstracts Service, Columbus, Ohio 43210

Received December 28, 1978

Chemical Abstracts Service provides a broad array of information and services for those who have an interest in fossil fuels. These services range from the time-honored CA Abstract Issues themselves, with their several types of indexes, through subsets of the printed abstracts selected and packaged according to various topical interests, to computer-readable tape files compiled for different, but complementary purposes.

Information on fossil fuels has been abstracted and indexed in *Chemical Abstracts* (CA) since its inception 70 years ago. The very first issue, January 1, 1907, carried 18 abstracts in a section entitled "Fuel, Gas, Coke", and 11 more were included in what was then an entirely separate section devoted to the patent literature. And in a section entitled "Petroleum, Asphalt, Wood Products", the only paper abstracted dealt with pitch and terpenes of pine and fir; the single pertinent patent was concerned with the refining of hydrocarbon oils.

In contrast, Issue 2 of Volume 89, dated July 1978, contained nearly 350 abstracts in the present Section 51, "Fossil Fuels, Derivatives, and Related Products". These numbers reflect not only the greatly increased interest in this field, but also the historic growth rate of the technical literature of some 8 to 9% per year.

THE ABSTRACTS

Chemical Abstracts Service (CAS) is charged with the responsibility of abstracting and indexing the chemical literature faithfully, as it is published, without exercising any critical evaluation as to its validity and/or intrinsic merit. Critical judgment is the responsibility of those who use CA. As distinct from certain other information services, CAS does not provide archival compilations of data; rather, it directs the user, through the abstracts and indexes (both for current work and retrospectively through the Collective Indexes), to the research papers, patents, conference proceedings, and books that report and discuss the scientific work carried out in the laboratories of the world. CA provides the access; others do the compiling.

Section 51 of the Abstract Issues of CA, "Fossil Fuels, Derivatives, and Related Products", is the principal section that deals with fossil fuels. The technical literature abstracted in this section encompasses the following aspects:[1]

(a) geochemistry: the origin, location, and description of deposits of petroleum, natural gas, oil shales, tar sands, coal, and peat;
(b) the chemistry and chemical engineering of extraction, production, processing, and usage;
(c) products and derivatives: synthetic fuel gases, gasoline, jet and diesel fuels, naphthas, lubricants and greases, bitumens and asphalt, coke, etc.;
(d) analysis of fossil fuels and their derivatives;
(e) organic compounds produced from carbon monoxide and hydrogen;
(f) synthetic lubricants and hydraulic fluids.

Explicitly excluded from this section are studies concerned primarily with the treatment of automotive exhausts and stack

† Presented, in part, before the Division of Chemical Information, Symposium on "Fossil Fuels Information Systems and Services", 174th National Meeting of the American Chemical Society, Chicago, Ill., Aug 31, 1977.

0095-2338/79/1619-0083$01.00/0